

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2004, 08:48:40 ; Search time 2824 Seconds
(without alignments)
10382.299 Million cell updates/sec

Title: US-10-822-496-5

Perfect score: 620
Sequence: 1 gtctcaaaataacatttca.....aaaaaaaaaaaaaaaaaaaaa 620

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 4526729 seqs, 23644849745 residues

Word size : 0

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: listing first 45 summaries

Database :

GenEmb1:
1: gb_ba:*
2: gb_hig:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	14	26.5	165	12 SCU87267	U87267 Synthetic c
2	29	4.7	77612	8 NCB8J24	AL669990 Neurospor
3	28	4.5	1845	5 BC044362	BC044362 Danio rer
4	28	4.5	2448	3 AF136344	AF136344 Toxoplas
5	28	4.5	204907	10 AL772249	AL772249 Mouse DNA
6	27	4.4	251	6 BD213305	BD213305 Novel hum
7	27	4.4	927	3 AMTRYPLP1	AMTRYPLP1
8	27	4.4	966	3 BC001872	BC001872 Homo sapi
9	27	4.4	966	3 BC004856	BC004856 Homo sapi
10	27	4.4	1453	9 AY130010	AY130010 Homo sapi
11	27	4.4	1629	10 BC049168	BC049168 Mus muscu
12	27	4.4	1705	5 BC047839	BC047839 Xenopus l
13	27	4.4	1992	5 BC044682	BC044682 Drosophi
14	27	4.4	2636	3 AY069866	AY069866 Drosophi
15	27	4.4	2905	3 AY099299	AY099299 Leishman
16	27	4.4	2916	3 AY099298	AY099298 Leishman
17	27	4.4	4548	3 AF414118	AF414118 Drosophi
18	27	4.4	101002	3 AY216936	AY216936 Plasmodi
19	27	4.4	101016	3 AY216939	AY216939 Plasmodi

20	27	4.4	101095	3 AY216937	AY216937 plasmodi
21	27	4.4	101158	3 AY216938	AY216938 plasmodi
22	27	4.4	160310	5 BX511089	BX511089 zebrafish
23	27	4.4	165777	10 BX004852	BX004852 mouse DNA
24	27	4.4	198057	2 AC073719	AC073719 Mus muscu
25	27	4.4	199866	2 AY003872	AY003872 Plasmodi
26	27	4.4	231443	2 AC073691	AC073691 Mus muscu
27	27	4.4	231524	2 AC095074	AC095074 Rattus no
28	27	4.4	253358	2 AC138303	AC138303 Mus muscu
29	26	4.2	157	6 CO466276	CO466276 Sequence
30	26	4.2	383	6 CO427072	CO427072 Sequence
31	26	4.2	416	3 AF299080	AF299080 Schistos
32	26	4.2	694	10 BC049612	BC049612 Mus muscu
33	26	4.2	793	6 BD019673	BD019673 Novel gen
34	26	4.2	793	6 BD098611	BD098611 Novel gen
35	26	4.2	983	6 AB035262	AB035262 Bipolaris
36	26	4.2	1054	6 AX214454	AX214454 Sequence
37	26	4.2	1054	6 AX951565	AX951565 Sequence
38	26	4.2	1185	3 AY094632	AY094632 Drosophi
39	26	4.2	1384	6 BD192727	BD192727 123 human
40	26	4.2	1384	6 AR184198	AR184198 Sequence
41	26	4.2	1696	6 AX685169	AX685169 Sequence
42	26	4.2	2066	3 AB072305	AB072305 Bombyx mo
43	26	4.2	2081	9 AB056423	AB056423 Macaca fa
44	26	4.2	2625	3 AY078995	AY078995 Brachios
45	26	4.2	2628	3 AY095026	AY095026 Drosophi

ALIGNMENTS

RESULT 1	SCU87267	165 bp	RNA	linear	SYN 05-MAR-1997
LOCUS	SCU87267				
DEFINITION	Synthetic construct promoting cardiac myofibrillogenesis in				
ACCESSION	U87267				
VERSION	U87267.1	GI:1842072			
KEYWORDS					
SOURCE	synthetic construct				
ORGANISM	artificial sequences.				
REFERENCE	1 (bases 1 to 165)				
AUTHORS	Lemanski, L.F., Nakatogawa, M., Bhatia, R., Erginel-Unaltuna, N., Spinner, B., and Dube, D.K.				
TITLE	A specific synthetic RNA promotes cardiac myofibrillogenesis in the Mexican axolotl				
JOURNAL	Mexican axolotl				
MEDLINE	97115686				
PUBMED	8955002				
REFERENCE	2 (bases 1 to 165)				
AUTHORS	Lemanski, L.F., Nakatogawa, M., Bhatia, R., Erginel-Unaltuna, N., Spinner, B., and Dube, D.K.				
TITLE	Direct Substitution				
JOURNAL	Submitted (27-JAN-1997) Anatomy and Cell Biology, SUNY-Health Science Center, 750 East Adams Street, Syracuse, NY 13210, USA				
FEATURES	Location/Qualifiers				
source	1..165				
	/organism="synthetic construct"				
	/mol_type="genomic RNA"				
	/db_xref="taxon:32630"				
	/lab_host="Ambystoma mexicanum"				
	/note="the synthetic construct was prepared using a sequence isolated from an axolotl cDNA library lab host strain=Lemanski"				
ORIGIN					
Query Match	26.5%; Score 164; DB 12; Length 165;				
Best Local Similarity	100.0%; Pred. No. 7.2e-87;				
Matches	164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
OY	312 AGACACATTCATTTTGGACACCTCTTACCGTGAAGAGGAGGAGCCGATCTT 371				

Tue Nov 23 09:27:45 2004

US-10-822-496-5.olin2p.rup

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: November 16, 2004, 13:21:52 / Search time 167.5 Seconds
(without alignments)
4259.490 Million cell updates/sec

Title: US-10-822-496-5

Perfect score: 201
Sequence: 1 gttcaaaataacatttaa.....aaaaaaaaaaaaaa 620

Scoring table:

OLIGO
Xgapop 60.0, Xgapext 60.0
Ygapop 60.0, Ygapext 60.0
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1825181 seqs, 575374646 residues

Word size: 1

Total number of hits satisfying chosen parameters: 3650028

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODE=frame_n2p.model -DEV=xip
-Q/cgn2 1/USPTO.spool.p/US10822496/runat.16112004.100156.24597/app.query.fasta_1.775
-DB=uniprot 02 -QFMT=fasta -SUFFIX=colindp.rup -MINMATCH=0.1 -LOOPT=0
-LOOEXT=0 -UNITS=bits -STRAP=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=6 -MAXLEN=2000000000
-USER=US10822496.gcg 1.1.244@runat.16112004.100156.24597 -NCP=6 -ICPU=3
-NO MAP -LANG=ENGLISH -NEG SCORES=0 -WAIT -DSPELCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database:

Uniprot 02: *
1: uniprot_sprot: *
2: uniprot_trembl: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
C 1	9	4.4	45	Q7RM43	Q7RM43 plasmodium
C 2	9	4.4	76	Q6Z201	Q6Z201 pyrochaculum
C 3	9	4.4	111	Y62C_YEAST	P53245 saccharomyc
C 4	9	4.5	120	Q7RLY6	Q7RLY6 plasmodium
C 5	9	4.5	130	Q9BUX1	Q9BUX1 plasmodium
C 6	9	4.4	150	Q95W70	Q95W70 trypanosoma
C 7	9	4.5	196	Q81J60	Q81J60 plasmodium
C 8	9	4.4	198	Q94DG3	Q94DG3 cryza sativ
C 9	9	4.5	214	Q81628	Q81628 plasmodium
C 10	9	4.5	221	Q7RID7	Q7RID7 plasmodium
C 11	9	4.5	240	Q6Q143	Q6Q143 pectus novy
C 12	9	4.5	240	KA86256	KA86256 ractus no
C 13	9	4.5	242	Q6P667	Q6P667 homo sapien
C 14	9	4.5	242	AA62433	AA62433 homo sapi
C 15	9	4.5	246	Q8CFP4	Q8CFP4 mus musculi
C 16	9	4.4	259	Q6JCU2	Q6JCU2 aleurodicus

17	9	4.5	280	Q9UF68	Q9UF68 homo sapien
18	9	4.5	280	Q9AWG7	Q9AWG7 cryza sativ
C 19	9	4.4	294	Q8KPI2	Q8KPI2 battonella
20	9	4.5	296	Q9WUW2	Q9WUW2 mus musculi
21	9	4.5	303	Q6PFA7	Q6PFA7 mus musculi
22	9	4.5	303	AA57655	AA57655 mus muscu
23	9	4.5	358	Q8XV92	Q8XV92 raltosia s
24	9	4.5	381	Q9H8P5	Q9H8P5 homo sapien
25	9	4.5	381	Q9BWK2	Q9BWK2 homo sapien
26	9	4.5	381	Q80UY2	Q80UY2 mus musculi
C 27	9	4.4	409	Q6PNA8	Q6PNA8 steinernema
C 28	9	4.4	409	AA00529	AA00529 steinernema
C 29	9	4.5	454	Q7RFH7	Q7RFH7 plasmodium
C 30	9	4.5	645	Q7R668	Q7R668 plasmodium
31	9	4.5	677	Q7RQ64	Q7RQ64 plasmodium
32	9	4.5	765	Q7RIN2	Q7RIN2 plasmodium
33	9	4.5	784	GCF_HUMAN	GCF_HUMAN
34	9	4.5	825	Q812K4	Q812K4 plasmodium
C 35	9	4.4	929	Q7RJ80	Q7RJ80 plasmodium
C 36	9	4.5	957	Q7RANH	Q7RANH plasmodium
37	9	4.5	1064	Q7RCM9	Q7RCM9 plasmodium
C 38	9	4.5	2010	Q96171	Q96171 plasmodium
39	9	4.5	2065	Q7RHY8	Q7RHY8 plasmodium
40	9	4.5	3604	Q7RRF7	Q7RRF7 plasmodium
41	9	4.5	4230	Q81C31	Q81C31 plasmodium
C 42	9	4.5	8094	Q81LB9	Q81LB9 plasmodium
C 43	8	3.9	18	Q8HKH6	Q8HKH6 apnomoma un
C 44	8	3.9	38	Q7RDE8	Q7RDE8 plasmodium
C 45	8	3.9	43	Q9HBA8	Q9HBA8 homo sapien

ALIGNMENTS

RESULT 1
ID Q7RM43 PRELIMINARY; PRT; 45 AA.
AC Q7RM43;
DT 01-MAR-2004 (TEMBLrel. 26, Created)
DT 01-MAR-2004 (TEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)
DE Hypothetical protein.
GN Name=PY02345;
OS Plasmodium yoelli yoelli.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNT;
RX PubMed=1236865;
RA Callion U.M., Angiuoli S.V., Suh B.B., Kooij T.W., Perlea M.,
RA Silva J.C., Ermolaeva W.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shalim S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabdi A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Valdivia A.B.,
RA van Lin L.H., Jansz C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.,
RT "Genome sequence and comparative analysis of the model rodent malaria
parasite Plasmodium yoelli yoelli.",
RL Nature 419:512-519 (2002).
CC -!- CAUTION: The sequence shown here is derived from an
EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL: ABL0100640; EAA21783.1; -
KM Hypothetical protein.
SQ SEQUENCE 45 AA; 5330 MM; 7C9638F4C9EAD77 CRC64;

Alignment Scores:
Pred. No.: 13.1 Length: 45
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0

US-10-822-496-5	(1-620)	x	Q7RM43	(1-45)
Best Local Similarity:	100.00%	Mismatches:	0	
Query Matchn:	4.41%	Indels:	0	
DB:	2	Gaps:	0	

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: November 16, 2004, 13:47:29 ; Search time 122.5 Seconds
(without alignments)
3581.516 Million cell updates/sec

Title: US-10-822-496-5
Perfect score: 201
Sequence: 1 gttcaaaatacatttca.....aaaaaaaaaaaaaaaa

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1568699 seqs, 35381937 residues

Word size: 1

Total number of hits satisfying chosen parameters: 3023662

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODE=frame+ n2p.model -DEV=xlp
-O/cgna2.1/uspro.spool.p/US10822496/runat.16112004.100159.24707/app.query.fasta_1.775
-DB=Published Applications AA -QWMT=faetan -SUFPRX=olin2p.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITG=bits -STAR=1 -END=1 -PATRX=olin2p.rapb -MINMATCH=0.1
-TRAN=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR_MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIDE=500 -MINLEN=0
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100
-LONGLOG -DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREDS=1 -XGAPOP=60 -XGAPEXT=60
-Fgapop=6 -Fgapext=7 -YGAPOP=60 -YGAEXT=60 -DELOP=6 -DELEXT=7

Database: Published Applications AA.*

1: /cgna2_6/ptodata/1/pubpa/US07_PUBCOMB.pep.*
2: /cgna2_6/ptodata/1/pubpa/PTC1_NEW_PUB.pep.*
3: /cgna2_6/ptodata/1/pubpa/US06_NEW_PUB.pep.*
4: /cgna2_6/ptodata/1/pubpa/US06_PUBCOMB.pep.*
5: /cgna2_6/ptodata/1/pubpa/US07_NEW_PUB.pep.*
6: /cgna2_6/ptodata/1/pubpa/PTCUS_PUBCOMB.pep.*
7: /cgna2_6/ptodata/1/pubpa/US08_NEW_PUB.pep.*
8: /cgna2_6/ptodata/1/pubpa/US08_PUBCOMB.pep.*
9: /cgna2_6/ptodata/1/pubpa/US09A_PUBCOMB.pep.*
10: /cgna2_6/ptodata/1/pubpa/US09B_PUBCOMB.pep.*
11: /cgna2_6/ptodata/1/pubpa/US09C_PUBCOMB.pep.*
12: /cgna2_6/ptodata/1/pubpa/US09_NEW_PUB.pep.*
13: /cgna2_6/ptodata/1/pubpa/US10A_PUBCOMB.pep.*
14: /cgna2_6/ptodata/1/pubpa/US10B_PUBCOMB.pep.*
15: /cgna2_6/ptodata/1/pubpa/US10C_PUBCOMB.pep.*
16: /cgna2_6/ptodata/1/pubpa/US10D_PUBCOMB.pep.*
17: /cgna2_6/ptodata/1/pubpa/US10_NEW_PUB.pep.*
18: /cgna2_6/ptodata/1/pubpa/US11_NEW_PUB.pep.*
19: /cgna2_6/ptodata/1/pubpa/US60_NEW_PUB.pep.*
20: /cgna2_6/ptodata/1/pubpa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query

No.	Score	Match	Length	DB	ID	Description
1	10	4.9	87	17	US-10-425-115-280203	Sequence 280203, App
2	10	5.0	656	17	US-10-739-930-5580	Sequence 5580, App
3	9	4.5	17	10	US-09-876-904A-345	Sequence 345, App
4	9	4.4	31	9	US-09-989-919-112	Sequence 112, App
5	9	4.5	37	17	US-10-425-115-208638	Sequence 208638, App
6	9	4.4	39	9	US-09-925-1515	Sequence 1515, App
7	9	4.4	39	10	US-09-925-299-1515	Sequence 1515, App
8	9	4.4	40	17	US-10-425-115-354630	Sequence 354630, App
9	9	4.5	41	16	US-10-437-963-174933	Sequence 174933, App
10	9	4.4	41	17	US-10-425-115-238266	Sequence 238266, App
11	9	4.4	42	17	US-10-425-115-309326	Sequence 309326, App
12	9	4.4	44	15	US-10-264-049-2354	Sequence 2354, App
13	9	4.4	44	15	US-10-424-599-152977	Sequence 152977, App
14	9	4.4	44	15	US-10-424-599-213738	Sequence 213738, App
15	9	4.4	47	16	US-10-437-963-163698	Sequence 163698, App
16	9	4.4	47	17	US-10-425-115-274699	Sequence 274699, App
17	9	4.4	50	16	US-10-437-963-182797	Sequence 182797, App
18	9	4.4	51	9	US-09-764-846-151	Sequence 151, App
19	9	4.4	51	14	US-10-091-483-151	Sequence 151, App
20	9	4.4	51	17	US-10-425-115-229065	Sequence 229065, App
21	9	4.4	52	15	US-10-424-599-163357	Sequence 163357, App
22	9	4.4	52	17	US-10-425-115-352124	Sequence 352124, App
23	9	4.4	53	15	US-10-424-599-274416	Sequence 274416, App
24	9	4.4	54	16	US-10-767-701-150915	Sequence 150915, App
25	9	4.4	54	17	US-10-425-115-319798	Sequence 319798, App
26	9	4.5	57	17	US-10-425-115-289672	Sequence 289672, App
27	9	4.4	58	14	US-09-764-846-183	Sequence 183, App
28	9	4.4	58	17	US-10-091-483-183	Sequence 183, App
29	9	4.4	58	17	US-10-425-115-188769	Sequence 188769, App
30	9	4.4	59	17	US-10-425-115-208268	Sequence 208268, App
31	9	4.4	59	17	US-10-425-115-269560	Sequence 269560, App
32	9	4.4	60	15	US-10-424-599-198167	Sequence 198167, App
33	9	4.4	60	17	US-10-425-115-265099	Sequence 265099, App
34	9	4.4	60	17	US-10-425-115-320384	Sequence 320384, App
35	9	4.4	61	9	US-09-764-846-195	Sequence 195, App
36	9	4.4	61	14	US-10-091-483-195	Sequence 195, App
37	9	4.4	61	15	US-10-424-599-239354	Sequence 239354, App
38	9	4.4	62	17	US-10-425-115-370566	Sequence 370566, App
39	9	4.4	63	15	US-10-424-599-149489	Sequence 149489, App
40	9	4.4	63	17	US-10-424-599-178837	Sequence 178837, App
41	9	4.4	63	15	US-10-425-115-335140	Sequence 335140, App
42	9	4.4	63	17	US-10-425-115-368838	Sequence 368838, App
43	9	4.4	64	15	US-10-424-599-177321	Sequence 177321, App
44	9	4.5	64	17	US-10-425-115-218215	Sequence 218215, App
45	9	4.5	64	17	US-10-425-115-356691	Sequence 356691, App

ALIGNMENTS

RESULT 1
US-10-425-115-280203
; Sequence 280203, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21 (3322)B
; CURRENT APPLICATION NUMBER: US/10/425, 115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 280203
; LENGTH: 87
; TYPE: PRT
; ORGANISM: Zea mays

OTHER INFORMATION: Clone ID: MRF4577_187121C.1.pep
US-10-425-115-280203

us-10-822-496-5.01in2p.rat

Run on: November 16, 2004, 13:33:28 ; Search time 28.5 Seconds

2885.415 Million cell updates/sec

Title: US-10-82-496-5
Perfect score: 201

Scoring table: OLIGO

Xgapop	60.0	Xgapext	60.0
Ygapop	60.0	Ygapext	60.0
Fgapop	6.0	Fgapext	7.0
Delop	6.0	Delext	7.0

Searched: 478139 seqs, 66318000 residues

word size

Total number of hits satisfying chosen parameters: 839080

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Listing first 45 summaries

Command line parameters

```

MODEL=ftime+az2.model -DBV-xlp
O-CG24.1/USTO.spool /psi0822496/runat.16.112004.100157.24629/app_query.faste.1.7755
DB-issued Patients AA -QWAT=fastan -SUPPLY=olimp.pai -MINMAX=human40.0.cdi
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.0.cdi
LIST=45 -DOCLINK=ext=000 -PRF SCORE=quality -TRR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTPM=pcio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=20000000000
-USER=SA10822496 @CSN.1.1.46 @runat.16.112004.100157.24629 -NCPU=6 -ICPU=3
NO_WAMP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCAL=100 -JONGLOG
DEVT-TIMEOUT=120 -WAMP-TIMEOUT=30 -THRAADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
FGAEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELXT=7

```

Database : Issued_Patents_AA:*

```

1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfill.esl.pep:

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	9	4.5	94	4	US-09-248-7964-22230	Sequence 22230, App
2	9	4.5	285	4	US-09-270-767-44439	Sequence 44439, App
3	8	4.0	13	4	US-09-325-601-11	Sequence 11, App
4	8	4.0	15	4	US-09-082-358B-24	Sequence 24, App
5	8	4.0	16	3	US-08-851-843X-201	Sequence 201, App
6	8	4.0	16	3	US-08-974-542X-320	Sequence 320, App
7	8	4.0	16	3	US-08-854-050-201	Sequence 201, App
8	8	4.0	16	3	US-09-430-323-201	Sequence 201, App
9	8	4.0	16	4	US-09-402-181B-320	Sequence 320, App
10	8	4.0	16	4	US-09-721-456-320	Sequence 320, App
11	8	3.9	17	4	US-09-148-545-221	Sequence 221, App
12	8	4.0	55	4	US-09-513-899C-6360	Sequence 6360, App

13	8	4.0	59	4	US-09-621-976-5183	Sequence 5183, A
14	8	4.0	60	4	US-09-248-7966-24364	Sequence 24364, A
15	8	3.9	66	4	US-09-248-7966-25981	Sequence 25981, A
16	8	3.9	67	4	US-09-270-767-74621	Sequence 74621, A
17	8	4.0	67	4	US-09-248-7966-24256	Sequence 24256, A
18	8	4.0	71	4	US-09-621-976-61184	Sequence 61184, A
19	8	4.0	74	4	US-09-621-976-61182	Sequence 61182, A
20	8	3.9	79	4	US-09-248-7966-27553	Sequence 27553, A
21	8	4.0	112	4	US-09-534-681A-83159	Sequence 83159, A
22	8	3.9	139	4	US-09-270-767-11067	Sequence 41067, A
23	8	3.9	139	4	US-09-270-767-56283	Sequence 56283, A
24	8	3.9	143	4	US-09-270-767-55811	Sequence 55811, A
25	8	3.9	143	4	US-09-270-767-51028	Sequence 51028, A
26	8	3.9	146	4	US-09-248-7966-21159	Sequence 21159, A
27	8	4.0	160	4	US-09-534-681A-8310	Sequence 8310, A
28	8	4.0	178	4	US-09-489-039A-8236	Sequence 8236, A
29	8	4.0	182	4	US-09-489-039A-8301	Sequence 8301, A
30	8	4.0	219	4	US-09-270-767-74502	Sequence 44502, A
31	8	4.0	246	4	US-09-538-092-1657	Sequence 1657, A
32	8	4.0	332	2	US-08-926-642-2	Sequence 2, A
33	8	4.0	332	3	US-08-778-354-4	Sequence 4, A
34	8	4.0	332	3	US-08-778-423A-2	Sequence 2, A
35	8	4.0	335	4	US-09-251-991A-26119	Sequence 26119, A
36	8	4.0	339	2	US-08-855-714-3	Sequence 3, A
37	8	4.0	345	4	US-09-248-7966-20831	Sequence 20831, A
38	8	4.0	376	4	US-09-328-352-1967	Sequence 1967, A
39	8	3.9	391	4	US-09-248-7966-15698	Sequence 15698, A
40	8	4.0	403	4	US-09-248-7966-19438	Sequence 19438, A
41	8	4.0	438	4	US-09-225-991A-20799	Sequence 20799, A
42	8	4.0	438	4	US-09-270-767-74168	Sequence 74168, A
43	8	3.9	438	4	US-09-248-7966-19853	Sequence 19853, A
44	8	3.9	480	4	US-09-100-123-1	Sequence 1, A
45	8	4.0	505	4	US-09-270-767-63185	Sequence 43185, A

ALIGNMENTS

RESULT 1

Sequence 22230, Application US/09248796A

; GENERAL INFORMATION:

1 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.13Z
CURRENT APPLICATION NUMBER:

CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725

PRIOR APPLICATION NUMBER: US 60/096,4

NUMBER OF SEQ ID NOS: 28208 ;

LENGTH: 94

ORGANISM: *Candida albicans*

Alignment Scores

Score:	9.00
Percent correct:	100

Best Local Similarity: 100.
Query Match: 4/49

DB: 4

US-10-822-496-5 (1-620) X US-09-248-796A-22230 (1-94

QY 594 AACGAGCAAAACAAAACAAAAA 820

cc
to

```

RESULT 2
US-09-270-767-44439
; Sequence 44439, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Hemdinger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44439
; LENGTH: 285
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-44439

Alignment Scores:
Pred. No.: 1.88 Length: 285
Percent Similarity: 9.00 Matches: 9
Conservative: 0 Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 4.48% Gaps: 0
DB: 4

US-10-822-496-5 (1-620) x US-09-270-767-44439 (1-285)
QY 594 AACGAGAAAAA 620
|||||
Db 190 ASnGtLysLysLysLysLysLys 198

RESULT 3
US-09-325-601-11
; Sequence 11, Application US/09325601
; Patent No. 6573645
; GENERAL INFORMATION:
; APPLICANT: Kazn
; TITLE OF INVENTION: Methods and kits for Discovery of RNA-Binding Compounds
; FILE REFERENCE: 3950/81235
; CURRENT APPLICATION NUMBER: US/09/325,601
; CURRENT FILING DATE: 1999-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Peptide from K,
US-09-325-601-11

Alignment Scores:
Pred. No.: 26.9 Length: 13
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.98% Indels: 0
DB: 4 Gaps: 0

US-10-822-496-5 (1-620) x US-09-325-601-11 (1-13)
QY 595 ACGAGAAAAA 618
|||||
Db 1 ThrArgLysLysLysLysLysLys 8

RESULT 4
US-09-082-358B-24
; Sequence 24, Application US/09082358B

```

```

? Patent No.6469153
? GENERAL INFORMATION:
? APPLICANT: Goff, Stephen P.
? APPLICANT: Li, Xingming
? TITLE OF INVENTION: EIP-1, EIP-3 GENES, ENVELOPE-INTERACTING PROTEINS
? TITLE OF INVENTION: EIP-1, and EIP-3
? FILE REFERENCE: 05/5/54804
? CURRENT APPLICATION NUMBER: US/09/082,356B
? CURRENT FILING DATE: 1998-05-20
? NUMBER OF SEQ ID NOS: 106
? SOFTWARE: PatentIn Ver. 2.1
? SEQ ID NO 24
? LENGTH: 15
? TYPE: PRT
? ORGANISM: murine
US-09-082-356B-24

Alignment Scores:
Pred. No.:          26.4          Length:          15
Score:              8.00          Matches:          8
Percent Similarity: 100.00%       Conservative:    0
Best Local Similarity: 100.00%    Mismatches:     0
Query Match:        3.98%         Indels:          0
DB:                 Gaps:         0

US-10-822-496-5 (1-620) x US-09-082-356B-24 (1-15)
Cy      597 GAGCAAAAAAAAAAAAAAAAAAAAAA 620
      |||||||
Db      5 Gtutysutysutysutysutys 12

RESULT 5
US-08-851-843A-201
? Sequence 201, Application US/08851843A
? Patent No. 6093809
? GENERAL INFORMATION:
? APPLICANT: Cech, Thomas R.
? APPLICANT: Lingner, Joachim
? APPLICANT: Nakamura, Tetsu
? APPLICANT: Chapman, Karen B.
? APPLICANT: Morin, Gregg B.
? APPLICANT: Harley, Calvin
? APPLICANT: Andrews, William H.
? TITLE OF INVENTION: No. 6093809e1 Telomerase
? NUMBER OF SEQUENCES: 225
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Townsend and Townsend and Crew LLP
? STREET: Two Embarcadero Center, 8th Floor
? City: San Francisco
? STATE: California
? COUNTRY: United States of America
? ZIP: 94111
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/951,843A
? FILING DATE: 06-MAY-1997
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/846,017
? FILING DATE: 25-APR-1997
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/844,419
? FILING DATE: 18-APR-1997
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/724,643
? FILING DATE: 01-OCT-1996
? CLASSIFICATION:

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: November 16, 2004, 13:20:52 ; Search time 125.5 Seconds

(without alignments)
3544.417 Million cell updates/sec

Title:

US-10-822-496-5

Perfect score:

201
1 gtcacaaataacattttaa.....aaaaaaaaaaaaaaaa 620

Sequence:

OLIGO

Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2002273 seqs, 358729299 residues

Word size:

1

Total number of hits satisfying chosen parameters: 3770738

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODE=frame+ n2p.model -DEV=xip
-Q=/cgr2/1/USPTO.spool/p/US10822496/runat.16112004.100155.24591/app.query.fasta_1.775
-DB=A.Geneseq.23Sep04 -QFMT=fasta -SUFFIX=colin2p.rag -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -STRAT=1 -END=1 -MATRIX=oligo -TRANS=human40.ctd
-LIST=45 -DOCALLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORV=ext -HEA=SIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10822496_QCGN_1_1_224_@runat.16112004.100155.24591 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -MAIN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -Fgapop=6
-Fgapext=7 -Ygapop=60 -Ygapext=60 -Delop=6 -Delext=7

Database :

A.Geneseq.23Sep04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	10	4.9	95	4	AA009153 Human pol
C 2	10	4.9	121	4	AA005271 Human pol
C 3	10	4.9	122	4	AA002187 Human pol
C 4	10	4.5	17	5	AB074581 Transcript
C 5	9	4.4	25	4	AA008699 Human pol
C 6	9	4.4	31	5	ABJ04338 Human col
C 7	9	4.4	39	3	AA053975 Human col
C 8	9	4.4	43	3	AA011858 Human col
C 9	9	4.4	44	5	ABP41222 Human ova
C 10	9	4.4	46	4	AA007812 Human pol

C 11	9	4.4	49	4	AA004514 Human pol
C 12	9	4.4	51	4	AA005079 Human pol
C 13	9	4.4	51	4	AA013671 Human pol
C 14	9	4.4	51	4	AA018166 Human pol
C 15	9	4.4	51	5	ABG92587 Human DNA
C 16	9	4.4	51	7	ADCC25304 Human ext
C 17	9	4.5	57	4	AA011486 Human pol
C 18	9	4.4	58	4	AA013804 Human pol
C 19	9	4.4	58	4	AA018198 Human pol
C 20	9	4.4	58	5	ABG92619 Human DNA
C 21	9	4.4	58	7	ADCC25316 Human ext
C 22	9	4.4	61	4	AA018210 Human pol
C 23	9	4.4	61	5	ABG92631 Human DNA
C 24	9	4.4	61	7	ADCC25348 Human ext
C 25	9	4.4	71	4	AA005367 Human pol
C 26	9	4.4	78	4	AA002152 Human pol
C 27	9	4.5	84	5	ABR01760 Human bre
C 28	9	4.4	87	4	AA013788 Human pol
C 29	9	4.4	88	4	AA003281 Human pol
C 30	9	4.4	89	4	AA007067 Human pol
C 31	9	4.4	91	4	AA006581 Human pol
C 32	9	4.4	92	5	ABP10718 Human ORF
C 33	9	4.4	102	4	AA007380 Human pol
C 34	9	4.4	102	4	AA010019 Human pol
C 35	9	4.4	103	4	AA005593 Human pol
C 36	9	4.4	104	4	AA007033 Human pol
C 37	9	4.4	106	3	AA018225 Human pol
C 38	9	4.4	106	3	AA007378 Human pol
C 39	9	4.4	106	7	ABO23529 Human pol
C 40	9	4.4	111	4	AA002315 Human pol
C 41	9	4.4	111	4	AA011818 Human pol
C 42	9	4.4	111	4	AA004871 Human pol
C 43	9	4.4	111	5	ABG93107 S. cerevi
C 44	9	4.4	112	4	AA003128 Human pol
C 45	9	4.4	114	4	AAU31171 Novel hum

ALIGNMENTS

RESULT 1
AA009153
ID AA009153 standard; protein, 95 AA.
AC AA009153;
XX
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 23045.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorders; arthritis; inflammation.
XX
OS Homo sapiens.
XX
XX WO200164835-A2.
XX
XX 07-SEP-2001.
XX
XX 26-FEB-2001; 2001WO-US004927.
XX
XX 28-FEB-2000; 2000US--00515126.
XX PR 18-MAY-2000; 2000US--00577409.
XX
XX (HISE) HISEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-514838/56.
XX DR N-PSDB; AA189884.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing

PT	and treating e.g. leukemia, inflammation and immune disorders.
XX	
PS	Claim 20; SEQ ID NO 23045; 1399pp + Sequence Listing; English.
XX	
CC	The invention relates to human polynucleotides (AA179941-AA193841) and
CC	the encoded proteins (AA00010-AA013910), that exhibit activity elating to
CC	cyclokin, cell proliferation or cell differentiation or which may induce
CC	production of other cytokines in other cell populations. The
CC	polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC	peptide therapy. The polypeptides have various cyclokin-like activities,
CC	e.g. stem cell growth factor activity, haematopoiesis regulating
CC	activity, tissue growth factor activity, immunomodulatory activity and
CC	activin/inhibin activity and may be useful in the diagnosis and/or
CC	treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC	inflammation. Note: The sequence data for this patent did not form part
CC	of the printed specification, but was obtained in electronic format
CC	directly from WIPO at ftp.wipo.int/pub/published_poc_sequences
XX	
SQ	Sequence 95 AA;
Alignment Scores:	
Pred. No.:	2.65 Length: 95
Score:	10.00 Matches: 10
Percent Similarity:	100.00% Conservatve: 0
Best Local Similarity:	100.00% Mismatches: 0
Query Match:	4.90% Indels: 0
Ds:	4 Gaps: 0
US-10-822-496-5 (1-620) x AA009153 (1-95)	
Oy	620 TTTTNTTTTTTTTTTTTTTTTCGTTGCA 591
Db	 78 PhephephephephephepheValArg 87
RESULT 2	
AA005271	
ID	AA005271 standard; protein; 121 AA.
XX	
AC	AA005271;
XX	
DT	06-NOV-2001 (first entry)
XX	
DE	Human polypeptide SEQ ID NO 19163.
XX	
KW	Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW	vaccine; peptide therapy; stem cell growth factor; haematopoesis;
KW	tissue growth factor; immunomodulatory; cancer; leukemia;
KW	nervous system disorders; arthritis; inflammation.
XX	
OS	Homo sapiens.
XX	
PN	WO200164835-A2.
PD	
XX	
XX	07-SEP-2001.
XX	
PF	26-FEB-2001; 2001WO-US004927.
XX	
PR	28-FEB-2000; 2000US-00515126.
ER	18-MAY-2000; 2000US-00577409.
XX	
FA	(HYSB-) HXSEO INC.
XX	
PI	Tang YT, Liu C, Dymnac RT;
XX	
WI	WPI; 2001-514838/56.
DR	N-PSDB; AAI85202.
XX	
PT	Isolated nucleic acids and polypeptides, useful for preventing diagnosing
XX	and treating e.g. leukemia, inflammation and immune disorders.
XX	
SQ	Claim 20; SEQ ID NO 19163; 1399pp + Sequence Listing; English.
CC	The invention relates to human polynucleotides (AA179941-AA193841) and

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: November 16, 2004, 12:58:46 / Search time 167 Seconds

(without alignments)
4272.243 Million cell updates/sec

Title: US-10-822-496-5
Perfect score: 1093
Sequence: 1 gtcataaatacatttca.....aaaaaaaaaaaaaaaa 620

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 3650362

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+np.model -DEV=xip
-Q=/cgn2_1/USPTO.spool_p/US10822496/runat_16112004_100129_24158/app_query.fasta-1.775
-DB=uniprot_02 -OPMT=fastan -SUFFIX=n2p.rup -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USR=US10822496 @CEN 1 1 244 @runat_16112004_100129_24158 -NCPU=6 -ICU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSELOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database: Uniprot_02:
1: uniprot_sprot:
2: uniprot_trembl:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No:	Score	Query Match	Length	DB ID	Description
C 1	95	8.3	725	2 Q9CV93	Q9CV93 mus musculus
C 2	94.5	8.6	315	2 Q6IGT5	Q6IGT5 drosophila
C 3	94	8.2	4782	2 Q8K1G6	Q8K1G6 mus musculus
C 4	91.5	8.0	256	2 Q96177	Q96177 plasmidium
C 5	91	8.0	5703	1 MUB8 HUMAN	Q9NC84 homo sapien
C 6	89.5	7.8	669	2 Q8N4X0	Q8N4X0 homo sapien
C 7	89.5	7.8	1070	2 Q96UG5	Q96UG5 homo sapien
C 8	88.5	7.7	3703	1 ABF1 HUMAN	Q15911 homo sapien
C 9	86.5	7.6	446	2 Q95ND3	Q95ND3 felis silve
C 10	86.5	7.6	609	2 Q959C3	Q959C3 galaxias po
C 11	86.5	7.6	684	2 Q97FM1	Q97FM1 oryza sativ
C 12	86.5	7.6	684	2 Q9M577	Q9M577 oryza sativ
C 13	86	7.5	700	2 Q8NF33	Q8NF33 homo sapien
C 14	84.5	7.4	579	2 Q7QH05	Q7QH05 anopheles g
C 15	84.5	7.7	782	1 L100 ADEL2	P36714 human adeno
C 16	84.5	7.7	782	2 Q7M6E4	Q7M6E4 human adeno

ALIGNMENTS

RESULT 1	ID	Q9CV93	PRELIMINARY;	PRT;	725 AA.
AC	Q9CV93	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)				
DT	01-MAR-2004 (TREMBLrel. 26, Last annotation update)				
DE	Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:230002104 product:similar to MUC1N (Fragment).				
GN	Name=Muc5b;				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C57BL/6J; TISSUE=Tongue;				
RX	MEDLINE=21085660; PubMed=11217851;				
RA	RIKEN FANTOM Consortium;				
RT	"Functional annotation of a full-length mouse cDNA collection.";				
RL	Nature 409:685-690(2001).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C57BL/6J; TISSUE=Tongue;				
RA	The RIKEN Genome Exploration Research Group Phase I & II Team;				
RT	"Analysis of the mouse transcriptome based on functional annotation of				
RL	60,770 full-length cDNAs.";				
RL	Nature 420:563-573(2002).				
RN	[4]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C57BL/6J; TISSUE=Tongue;				
RX	MEDLINE=20499374; PubMed=11042159;				

17	84.5	7.7	990	2 Q6NR11	Q6NR11 drosophila
18	84.5	7.7	990	2 AAQ23591	AAQ23591 drosophila
19	84.5	7.7	1052	2 Q9VF12	Q9VF12 drosophila
20	84.5	7.7	1138	2 Q9VF13	Q9VF13 drosophila
C 21	84	7.3	381	2 Q915P2	Q915P2 pseudomona
C 22	84	7.3	2848	2 Q6W705	Q6W705 cryphonect
C 23	84	7.3	2848	2 AAQ76546	AAQ76546 cryphonect
C 24	83.5	7.3	404	2 Q6ZOR7	Q6ZOR7 homo sapien
C 25	83.5	7.3	404	2 BAC87616	BAC87616 homo sapi
C 26	83.5	7.3	611	2 Q959A5	Q959A5 galaxias an
C 27	83.5	7.6	824	2 Q6C200	Q6C200 yarrowia ii
C 28	83	7.3	1707	2 Q6P9K2	Q6P9K2 mus musculu
C 29	83	7.3	1707	2 AAH60729	AAH60729 mus muscu
C 30	83	7.3	2192	2 Q01768	Q01768 caenorhabdi
C 31	83	7.3	3726	1 ABF1 MOUSE	ABF1 MOUSE
C 32	82.5	7.5	233	2 Q9UFP2	Q9UFP2 mus musculu
C 33	82.5	7.5	289	2 Q9NUN6	Q9NUN6 homo sapien
C 34	82.5	7.5	385	2 Q86VC5	Q86VC5 homo sapien
C 35	82.5	7.5	533	2 Q9H8Z2	Q9H8Z2 homo sapien
C 36	82.5	7.5	584	2 Q6IN54	Q6IN54 homo sapien
C 37	82.5	7.5	584	2 AAH72453	AAH72453 homo sapi
C 38	82.5	7.2	611	2 Q959B3	Q959B3 galaxias br
C 39	82.5	7.2	1740	2 Q6ENK5	Q6ENK5 oryza sativ
C 40	82.5	7.5	2312	1 CCAB_HUMAN	CCAB_HUMAN
C 41	82	7.2	325	1 HMGJ_RAT	HMGJ_RAT
C 42	82	7.2	325	2 AAH61797	AAH61797 rattus norv
C 43	82	7.2	358	2 Q6N172	Q6N172 corynebacte
C 44	82	7.2	358	2 CAE49422	CAE49422 corynebac
C 45	82	7.5	625	2 P93520	P93520 zea mays (m

RA Carninci P., Shibata Y., Hayatsu N., Sugata Y., Shibata K., Itoh M.,
RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.,
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.",
RL Genome Res. 10:1617-1630(2000).
RN [5]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RC MEDLINE=2053091; Pubmed=11076861;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Kono H., Akiyama J., Nishi K., Katsunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishi Y., Nakamura S., Hahama M., Nishino T., Harada A.,
RA Yamamoto R., Matsunoto H., Sakauchi S., Ikegami T., Kashiwagi K.,
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanabe M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 984 multicapillary sequencer.",
RL Genome Res. 10:1757-1771(2000).
RN [6]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA Arakawa T., Bono H., Carninci P., Fukuda S., Furumoto K., Furuo M.,
RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hirooka T., Hori F.,
RA Imotani K., Ishi Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA Kawai J., Kojima Y., Kono H., Konda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ono M.,
RA Okazaki Y., Okino T., Owa C., Saito K., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shikigawa A., Shiraki T.,
RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Taniguchi F., Tanaka T.,
RA Teijima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.,
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK009011; BAB26024.1; -
DR MGD; MGI:1921430; Muc5b.
DR GO; GO:00055615; C:extracellular space; IEA.
DR GO; GO:0005386; F:carrier activity; IEA.
DR GO; GO:0006810; P:transport, IEA.
DR InterPro; IPR002915; Cysrich_T1L.
DR InterPro; IPR006207; Cys_knot_C.
DR InterPro; IPR005041; PMF_SGCT.
DR InterPro; IPR0010264; Serum_albumin.
DR InterPro; IPR001007; VWF_C.
DR InterPro; IPR001846; VWF_D.
DR Pfam; PF00093; VMC; 1.
DR Pfam; PF00094; VMD; 1.
DR SMART; SM00641; CT; 1.
DR SMART; SM00214; VMC; 2.
DR SMART; SM00216; VMD; 1.
DR PROSITE; PS00212; ALBUMIN; 1.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01225; CTCK_2; 1.
DR PROSITE; PS01208; VWF_C; 1.
DR PROSITE; PS01018; UNKNOWN_2.
DR PROSITE; PS01018; VWF_C; 2.
FT NON TER 1
FT 1
SQ SEQUENCE 725 AA; 78872 MW; F2D96B2CAB755264 CRC64;

Alignment Scores:

Pred. No.:	2.2	Length:	725
Score:	95.00	Matches:	62
Percent Similarity:	29.66%	Conservative:	24
Best Local Similarity:	21.38%	Mismatches:	67
Query Match:	8.30%	Indels:	137
DB:	2	Gaps:	16

US-10-832-496-5 (1-620) x Q9CV93 (1-725)

QY AGTCCGCTTATCTATCGCCGCCCCCGCTTATCTGCGCTTAT-----529

Db 229 SeiserSersleu-AlaGlnProserSero-----LeuProserThrThrProGlnThrPr 246

528 -GTTGCCCTTTCTCTACTTCCTCCGCGCCCTCCGCTCCATATAGCTGCCACATGTCT 470

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: November 16, 2004, 13:06:41 ; Search time 36 Seconds

(without alignments)
3314.133 Million cell updates/sec

Title: US-10-822-496-5

Perfect score: 1093

Sequence: 1 gttcaataataacatttca.....aaaaaaaaaaaaaaaaaaaaa 620

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Xgapop 10.0 , Xgapext 0.5
Xgapop 6.0 , Xgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283416 segs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL-frame.n2p.model -DEV=klp
-Q/cg21/USFTO_spool_p/US10822496/runat.16112004.100130.24178/app.query.fasta.1.775
-DB=PIR_79 -QMT=faстан -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFN=pcio -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USRN=US10822496 @CGN 1.1 63 @runat.16112004.100130.24178 -NCPU=6 -ICPU=3
-NO MMP -LARGEJOB -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEY TIMEOUT=120 -MAPN TIMEOUT=30 -THRPADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

PIR_79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	88.5	7.7	2783	1 A41948	alpha-fetoprotein
C 2	84.5	7.7	782	2 S33945	late protein, 100K
C 3	84	7.5	381	2 A83561	probable type II s
C 4	82.5	7.5	233	2 T17218	hypothetical prote
C 5	82	7.5	625	2 T02033	calcium/calmodulin
C 6	81.5	7.1	391	1 S15617	E2 protein - human
C 7	81.5	7.1	496	2 T52132	probable sugar tra
C 8	81	7.1	427	2 S60742	alpha-transinducin
C 9	81	7.1	427	2 JC2203	U48h protein - Ma
C 10	81	7.1	3014	1 UC5620	genome polyprotein
C 11	80.5	7.0	587	2 H72647	hypothetical prote
C 12	80	6.9	587	2 T16867	probable cytochrom
C 13	79.5	6.9	354	2 T22274	hypothetical prote
C 14	79.5	7.3	1199	2 T15826	hypothetical prote

C 15	79	7.2	385	2 H89046	protein C10G8.8 (i
C 16	79	6.9	388	2 S36500	E2 protein - human
C 17	77	6.7	620	2 H69382	ABC transporter, A
C 18	76.5	6.7	655	2 T16538	hypothetical prote
C 19	76.5	6.7	730	2 A96788	protein T4012.13 (
C 20	76	6.6	321	2 P84597	probable prolone-r
C 21	76	6.6	329	2 B72380	m4C-methyltransfer
C 22	76	7.0	328	2 A72048	proteinkinase (impo
C 23	76	6.6	1172	1 TSHD2	chromospondin 2 p
C 24	75.5	6.6	347	2 A75483	hypothetical prote
C 25	75.5	6.9	348	2 S57601	ribonuclease H (EC
C 26	75.5	6.6	402	2 S11773	polygalacturonase
C 27	75.5	6.6	1459	2 A12468	hypothetical prote
C 28	75	6.9	168	2 T40352	single-stranded DN
C 29	75	6.6	1056	2 A53787	mucln MUC5B, trach
C 30	75	6.6	1194	2 B36624	hypothetical prote
C 31	74.5	6.5	213	2 F84532	hypothetical prote
C 32	74.5	6.5	378	2 T46268	brevican precursor
C 33	74.5	6.5	515	2 S53525	G protein-coupled
C 34	74.5	6.5	1105	2 T18295	Ap-3 adaptor compl
C 35	74.5	6.8	2288	2 T29999	hypothetical prote
C 36	74.5	6.8	2477	2 S14428	fibronectin precu
C 37	74	6.8	298	2 JCT870	ribonuclease Ie3 -
C 38	74	6.5	325	2 I57009	3-Hydroxy-3-methyl
C 39	74	6.5	340	2 B22360	Ig alpha-2 chain C
C 40	74	6.5	750	2 T42614	probable envelope
C 41	74	6.5	1246	2 C36806	hypothetical prote
C 42	74	6.5	1265	2 S17675	valine-tRNA ligase
C 43	74	6.5	1496	2 T40016	conserved hypothet
C 44	74	6.5	2325	2 A61208	chondroitin sulfat
C 45	73.5	6.4	262	2 T14610	hypothetical prote

ALIGNMENTS

RESULT 1
A41948
alpha-fetoprotein enhancer-binding protein - human
N:Alternate names: A7BFL protein
C:Species: Homo sapiens (man)
C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 15-Oct-1999
C/Accession: A41948
C/Accession: T.; Yasuda, H.; Hashimoto, T.; Higashio, K.; Tamaoki, T.
Mol. Cell. Biol. 11, 6041-6049, 1991
A>Title: A human alpha-fetoprotein enhancer-binding protein, A7BFL, contains four homeod
A:Reference number: A41948; MUID:92049333; PMID:1719379
A:Accession: A41948
A:Molecule type: mRNA
A:Residues: 12783 <MOR>
A:Cross-references: GB:D10250; GB:D90395; NTD:9219429; PTDN:BA01095.1; PTD:G219430
A/Note: sequence extracted from NCBI backbone (NCBI:66271, NCBI:66276)
C/Genetics:
A:Gene: GDB:A7BFL
A:Cross-references: GDB:392090; OMIM:104155
A/Map position: 16q22.3-16q23.1
C:Superfamily: alpha-fetoprotein enhancer-binding protein; homeobox homology
C/Keywords: DNA binding; homeobox; nucleus; transcription regulation; zinc finger
P:12.94/Region: zinc finger CCH motif
F:128-150/Region: zinc finger CCH motif
F:176-198/Region: zinc finger CCH motif
F:311-332/Region: zinc finger CCH motif
F:340-361/Region: zinc finger CCH motif
F:448-471/Region: zinc finger CCH motif
F:489-509/Region: zinc finger CCH motif
F:517-538/Region: zinc finger CCH motif
F:633-655/Region: zinc finger CCH motif
F:684-706/Region: zinc finger CCH motif
F:719-773/Region: serine/threonine-rich
F:809-958/Region: glutamine-rich
F:1071-1092/Region: zinc finger CCH motif
F:1117-1211/Region: proline-rich
F:1232-1288/Domain: homeobox homology <HOX1>
F:1329-1385/Domain: homeobox homology <HOX2>

Tue Nov 23 09:27:40 2004

us-10-822-496-5.n2p.rapb

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: November 16, 2004, 13:19:42 : Search time 122 Seconds
(without alignments)
3596.195 Million cell updates/sec

Title: US-10-822-496-5
Perfect score: 1093
Sequence: 1 gttcaataatcattttaa.....aaaaaaaaaaaaaa 620

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 156869 segs, 353819137 residues

Total number of hits satisfying chosen parameters: 3137398

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame-n2p-model -DEV=x1p
-Q/cgnt1/usfto.spool.p/us10822496/runatc_16112004_100132_24268/asp_query.fasta.1.775
-DB=Published.Applications_AA -QWMT=faastaan -SUFFIX=n2p.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=10sum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=ext -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=plc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10822496 @CGN 1.1.199 @runatc_16112004_100132_24268
-NCU=6 -ICPU=3 -NO MMAP -IARQUEURY -NEG SCORES=0 -WAIT -DSPELOCK=100
-LONGLOG -DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXIT=7

Database : Published.Applications_AA:

1: /cgnt2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgnt2_6/ptodata/1/pubpaa/PCF_NEW_PUB.pep:*
3: /cgnt2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgnt2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgnt2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgnt2_6/ptodata/1/pubpaa/PCF_US07_PUBCOMB.pep:*
7: /cgnt2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgnt2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgnt2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
10: /cgnt2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgnt2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgnt2_6/ptodata/1/pubpaa/US09C_NEW_PUB.pep:*
13: /cgnt2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
14: /cgnt2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
15: /cgnt2_6/ptodata/1/pubpaa/US10C_NEW_PUB.pep:*
16: /cgnt2_6/ptodata/1/pubpaa/US10C_NEW_PUB.pep:*
17: /cgnt2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep:*
18: /cgnt2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep:*
19: /cgnt2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
20: /cgnt2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result Query

No.	Score	Match	Length	DB	ID	Description
C 1	117.5	10.3	327	14	US-10-017-161-2258	Sequence 2258, Ap
C 2	117.5	10.3	327	14	US-10-292-798-1904	Sequence 1904, Ap
C 3	99	9.1	150	16	US-10-767-701-45440	Sequence 45440, A
C 4	92	8.0	153	16	US-10-437-963-156345	Sequence 156345, A
C 5	91.5	8.0	137	15	US-10-424-599-145989	Sequence 145989, A
C 6	89	7.8	159	17	US-10-423-115-23612	Sequence 23612, A
C 7	88.5	8.1	201	16	US-10-437-963-149326	Sequence 149326, A
C 8	88.5	8.1	426	16	US-10-767-701-44661	Sequence 44661, A
C 9	88.5	7.7	2783	9	US-09-816-669A-14	Sequence 14, Appl
C 10	88.5	7.7	2789	16	US-10-408-765A-2228	Sequence 2228, Ap
C 11	88	7.7	124	16	US-10-437-963-177169	Sequence 177169, A
C 12	87.5	7.6	178	17	US-10-425-115-241376	Sequence 241376, A
C 13	87.5	7.6	189	17	US-10-425-115-231208	Sequence 231208, A
C 14	86.5	7.6	102	17	US-10-425-115-230642	Sequence 230642, A
C 15	86.5	7.6	901	16	US-10-437-963-170537	Sequence 170537, A
C 16	86	7.5	106	15	US-10-424-599-23677	Sequence 23677, A
C 17	86	7.5	657	15	US-10-294-006-13	Sequence 13, Appl
C 18	86	7.5	895	15	US-10-294-006-35	Sequence 35, Appl
C 19	85	7.4	98	15	US-10-424-599-260338	Sequence 260338, A
C 20	85	7.4	203	15	US-10-426-774-2002	Sequence 2002, Ap
C 21	84.5	7.4	132	17	US-10-425-115-219535	Sequence 219535, A
C 22	84.5	7.4	286	17	US-10-425-115-330494	Sequence 330494, A
C 23	84	7.3	111	16	US-10-437-963-204029	Sequence 204029, A
C 24	84	7.3	160	16	US-10-437-963-188033	Sequence 188033, A
C 25	83.5	7.3	246	14	US-10-017-161-2056	Sequence 2056, Ap
C 26	83.5	7.3	246	14	US-10-292-798-1702	Sequence 1702, Ap
C 27	83.5	7.3	274	14	US-10-128-714-3124	Sequence 3124, Ap
C 28	83.5	7.3	274	14	US-10-128-714-3124	Sequence 3124, Ap
C 29	83.5	7.6	341	14	US-10-091-438-166	Sequence 166, App
C 30	83.5	7.3	1057	14	US-10-189-971-6	Sequence 6, Appl
C 31	83.5	7.3	1192	14	US-10-189-971-18	Sequence 18, Appl
C 32	83.5	7.3	1251	14	US-10-189-971-20	Sequence 20, Appl
C 33	83.5	7.3	1432	14	US-10-189-971-24	Sequence 24, Appl
C 34	83.5	7.3	1477	14	US-10-189-971-28	Sequence 28, Appl
C 35	83.5	7.3	1512	14	US-10-189-971-10	Sequence 10, Appl
C 36	83.5	7.3	1535	14	US-10-189-971-14	Sequence 14, Appl
C 37	83.5	7.3	1570	14	US-10-189-971-12	Sequence 12, Appl
C 38	83.5	7.3	1593	14	US-10-189-971-4	Sequence 4, Appl
C 39	83.5	7.3	1628	14	US-10-189-971-2	Sequence 2, Appl
C 40	83.5	7.3	253	15	US-10-424-599-157200	Sequence 157200, A
C 41	83	7.6	77	15	US-10-424-599-235133	Sequence 235133, A
C 42	82.5	7.5	143	17	US-10-425-115-344539	Sequence 344539, A
C 43	82.5	7.2	533	15	US-10-051-874-137	Sequence 137, App
C 44	82.5	7.5	533	15	US-10-408-765A-1052	Sequence 1052, Ap
C 45	82.5	7.5	533	16	US-10-408-765A-1052	Sequence 1052, Ap

ALIGNMENTS

RESULT 1
US-10-017-161-2258
; Sequence 2258, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUMA, YAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABEURANT, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; PRIOR APPLICATION NUMBER: 2002-12-18
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2258
; LENGTH: 327
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: November 16, 2004, 12:58:01, Search time 125 Seconds

(without alignments)
3558.595 Million cell updates/sec

Title: US-10-822-496-5

Perfect score: 1093
Sequence: 1 gttcaataataacattttaa.....aaaaaaaaaaaaaaaaaaaa 620

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Zgapop 6.0, Zgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 4004546

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p_model -DEV=r1p
-Q=/cgrt_1/USP70.spool_p/US10822496/runat_16112004_100129_24151/app_query.fasta_1.775
-DB=us-10-822-496-5 -QFMT=fastan -SUFFIX=n2p_rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10822496 @CGN 1.1 224 @runat 16112004 100129 24151 -NCPU=6 -ICPU=3
-NO MMAP -LARGEOUTRY -NEG_SCORES=0 -AIR -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -RAPOP=6
-FgapExt=7 -Ygapop=10 -YgapExt=0.5 -DELOP=6 -DELEXT=7

Database: A_Geneseq_23Sep04:*

1: Geneseq1980s:*\n2: Geneseq1990s:*\n3: Geneseq2000s:*\n4: Geneseq2001s:*\n5: Geneseq2002s:*\n6: Geneseq2003as:*\n7: Geneseq2003bs:*\n8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	117.5	10.3	327	7	ADC87451 Human GPC
C 2	91	8.0	4315	5	ABP43908 MUC5B par
C 3	91	8.0	5703	8	AD123265 Human MUC
C 4	90.5	8.3	112	7	ADFO8034 Bacterial
C 5	89.5	7.8	1030	6	ABRS8606 Human can
C 6	89.5	7.8	1070	7	ADFS5479 Human nov
C 7	88.5	7.7	2763	2	AAR23963 Human pro
C 8	88.5	7.7	2763	2	AAR23962 AFP-1 (AI
C 9	88.5	7.7	2763	4	AAB82946 Human and
C 10	88.5	7.7	2763	7	ADJ70422 Human hea

11	87	8.0	171	7	ABO61718 Klebsiell
C 12	86	7.5	657	6	ABG70519 Human pol
C 13	86	7.5	657	8	ADJ92551 Human pro
C 14	86	7.5	895	8	ADJ92573 Human pro
C 15	85.5	7.5	113	4	ABG09348 Novel hum
C 16	85.5	7.5	718	4	ABG04957 Novel hum
C 17	85	7.4	169	7	ABO68122 Pseudom
C 18	85	7.4	203	4	ABE11632 Human dop
C 19	84.5	7.4	145	4	ABG28355 Novel hum
C 20	84.5	7.4	1058	4	ABE71401 Drosophil
C 21	83.5	7.3	246	7	ADC87249 Human GPC
C 22	83.5	7.3	274	6	ABJ26066 Aspergill
C 23	83.5	7.3	274	6	ABJ25466 Aspergill
C 24	83.5	7.6	341	4	AAU18021 Human imm
C 25	83.5	7.6	341	7	ADB31645 Human nov
C 26	83.5	7.3	1057	6	ABP97370 Human Xie
C 27	83.5	7.3	1192	6	ABP97376 Human Xie
C 28	83.5	7.3	1207	6	ABP97377 Human Xie
C 29	83.5	7.3	1251	6	ABP97375 Human Xie
C 30	83.5	7.3	1342	6	ABP97379 Human Xie
C 31	83.5	7.3	1477	6	ABP97371 Human Xie
C 32	83.5	7.3	1512	6	ABP97372 Human Xie
C 33	83.5	7.3	1515	6	ABP97374 Human Xie
C 34	83.5	7.3	1570	6	ABP97373 Human Xie
C 35	83.5	7.3	1593	6	ABP97369 Human Xie
C 36	83.5	7.3	1628	6	ABP97368 Human Xie
C 37	83	7.3	153	4	AAU31783 Novel hum
C 38	82.5	7.2	160	4	ABG27888 Novel hum
C 39	82.5	7.5	178	7	ABO61719 Klebsiell
C 40	82.5	7.2	258	7	ABO71079 Pseudom
C 41	82.5	7.5	289	4	AAB93619 Human pro
C 42	82.5	7.2	291	6	ABR01685 Human pol
C 43	82.5	7.5	519	4	AAM04360 Human cel
C 44	82.5	7.5	533	4	AAB60463 Human cel
C 45	82.5	7.5	533	4	AAB95321 Human pro

RESULT 1	ADC87451	standard: protein: 327 AA.	ALIGNMENTS
XX	ADC87451;		
XX	01-JAN-2004 (first entry)		
XX	Human GPCR protein SEQ ID NO:1904.		
XX	human; GPCR; guanosine triphosphate-binding protein coupled receptor;		
XX	gene therapy.		
XX	Homo sapiens.		
XX	EP1270724-A2.		
XX	02-JAN-2003.		
XX	18-JUN-2002; 2002EP-00013517.		
XX	18-JUN-2001; 2001JP-00246789.		
XX	(NAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY INCUBATIO.		
XX	(ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.		
XX	Suwa M, Arai K, Akiyama Y, Aburatani H;		
XX	WPI: 2003-315783/31.		
XX	N-PSDB; ADC87450.		
XX	New polynucleotide, useful for preparing a composition for treating a		
XX	patient in need of increased or suppressed activity or expression of the		
XX	guanosine triphosphate-binding protein coupled receptor.		

[illegible]

	XX	New polypeptides and their encoded proteins, useful as nutritional
	PT	sources or supplements, or in gene therapy, particularly for treating
	PT	wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or
	PT	inflammation.
	XX	
	PS	Claim 20; SEQ ID # 811; 357bp + Sequence Listing; English.
	XX	
	CC	The invention relates to 446 newly isolated polynucleotide sequences. The
	CC	activity of polynucleotides of the invention may be described as,
	CC	vulnerable, neuroprotective, immunomodulator, cytostatic and anti-
	CC	inflammatory. Compositions comprising nucleic acids of the invention are
	CC	useful for treating a mammalian subject, or as nutritional sources or
	CC	supplements. These are useful in gene therapy, particularly for treating
	CC	wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,
	CC	amyotrophic lateral sclerosis, autoimmune disorders, cancer or
	CC	inflammation. The nucleic acids and polypeptides are also useful in
	CC	diagnostic and research methods. The sequences given in records ABP43544-
	CC	ABP4389 represent polypeptides encoded by polynucleotides of the
	CC	invention. NOTE: The sequence data for this patent did not form part of
	CC	the printed specification, but was obtained in electronic format directly
	CC	from WIPD at ftp.wipd.int/pub/published_pct_sequences
	XX	
	SQ	Sequence 4315 AA;
	Alignment Scores:	
	Pred. No.: 1.06 Length: 4315	
	Score: 91.00 Matches: 52	
	Percent Similarity: 34.56% Conservative: 23	
	Best Local Similarity: 23.96% Mismatches: 60	
	Query Match: 7.95% Indels: 82	
	Ds: 5 Gaps: 11	
	US-10-822-496-5 (1-620) x ABP43908 (1-4315)	
OY	ACGTTTCTCCACAGCAGTCCTGGTTATTCTCGTGCCCTCCGCCCTTTATCTGCCTCT	532
DB	3822 ThAlaserPro--AlAlalProvalSerSerThPrOthnPro-----	3835
OY	531 ATTGTTGCCCTTCTCTCACTCCGCCCTGCCCTCCCATACAGTGTGCCAATTGT	472
DB	3836 -----ThPrOcysrProGlInrProleucysAspLeuMet	3847
OY	471 CTTTCTGATCTCCATTCGAAAGCCGATAGAAGAGTSCCT-----GCTTAATCCAGCAA	418
DB	3848 LeuSerGIlnValRheAlGIlnCushIAsnIleValrProGrOIyProRhehemaIA	3867
OY	417 CTGCTGGATATGCGCTGC-----AACCTGAGATGCATGATACAATTCGAAAGATCG	364
DB	3868 CysIIseSerAsnHIScysArgelnyArgIenGIlnValrProCysdlSenIeudlunIAr	3887
OY	363 GCTCCGCCCTTCATGCCAGGTAGAGAGAGGTCTCCAAAATGAAGTGGTCTCACCGT	304
DB	3888 Ala---GIlnIeCysatgAlaIargIlyValCysSerzArPrIdArgelnyAlathr---	3904
OY	303 GTCCAAAGGATACCGGCTCGGACATGTTGAAGCTCGAGTGGTGAAG-----	257
DB	3905 -GIlyIeucCysAspLeuThrCysrProProthrInIlyValrlyrProCysglProI	3924
OY	256 -----GCTCAACGAAATGAAGTTTCGCGTATTGGCGTTC	223
DB	3924 IeelnrProAlathrCysAsnSerIrgAsnIeetrProclIneugIueIymetAlagIUG	3944
OY	222 ATGTGTTTCTTGGTCAGTCAGTTGGAAGATATATTCACATAACATGGAAAACTTGCA	163
DB	3944 IyCysrPhe-----	3946
OY	162 CCTATTACATTGCGCTAGAGAAAGTGCCTGTCTC-----	125
DB	3947 -----CysrProdlAsprlnIleIeurhemaSnIAhIsmetGIlyIeCysValG	3963
OY	124 -----CCGGCTATC-----CAGCACTTCCGAA-----	101
DB	3963 InAlAcysrCysValGIyProArGrOIyRheProCylusPheProGIyUatIGTrVAIs	3993

Tue Nov 23 09:27:43 2004

us-10-822-496-5.01g.rst

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2004, 08:42:52 ; Search time 2993 Seconds
(without alignments)
7548.491 Million cell updates/sec

Title: US-10-822-496-5

Perfect score: 620

Sequence: 1 gttcaaaatacatttca.....aaaaaaaaaaaaaaaaaaaaa 620

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 32822875 seqs, 1821985908 residues

Word size : 0

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*

1: gb_est1:.*
2: gb_est2:.*
3: gb_hic:.*
4: gb_est3:.*
5: gb_est4:.*
6: gb_est5:.*
7: gb_est6:.*
8: gb_gsal:.*
9: gb_gsal2:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	4.7	543	7	CK360119 AGENCOURT
2	28	4.7	601	2	BF681816 602117624
3	28	4.5	199	1	A1365051 q12b04.x
4	28	4.5	285	2	BE049016 xre4f11.x
5	28	4.5	355	4	BM378616 MEST567-F
6	28	4.5	384	7	CN171083 AGENCOURT
7	28	4.5	406	4	BI510807 B160004A
8	28	4.5	585	6	CD666951 NCST3B93
9	28	4.5	590	5	BM074026 MEST79-B1
10	28	4.5	621	4	CF515989 CAP0002.1
11	28	4.5	676	3	CR695006 Terradon
12	28	4.5	910	3	CR673159 Terradon
13	28	4.5	921	3	CR673159 Terradon
14	28	4.5	949	7	CO757970 brain.EST
15	28	4.5	1101	9	AI083694 Drosophila
16	28	4.5	1104	3	CR679551 Terradon
17	28	4.4	99	7	CK736897 T9851yko
18	27	4.4	165	4	BI813177 J005B08.O
19	27	4.4	171	1	A1105935 b03h08.P
20	27	4.4	186	6	CD722635 o12b06.s1
21	27	4.4	211	7	N80246 za96b01.s1
22	27	4.4	212	5	BU531282 AGENCOURT
23	27	4.4	214	6	CD724006 c72b03.y
24	27	4.4	236	2	AM073961 x05f06.x

25	27	4.4	236	4	BI512339	BI512339 BB160008B
26	27	4.4	250	7	CK678589	CK678589 2F101-P00
27	27	4.4	274	2	BE530231	BE530231 602071406
28	27	4.4	276	6	CD673937	CD673937 f805d07.x
29	27	4.4	292	6	CB709984	CB709984 AMGNNUC.U
30	27	4.4	293	6	CF202144	CF202144 KR890915N
31	27	4.4	308	6	CB048748	CB048748 NISC_9106
32	27	4.4	314	5	BU565818	BU565818 AGENCOURT
33	27	4.4	342	2	AM446955	AM446955 87884.MAR
34	27	4.4	349	1	AI022895	AI022895 0544f10.s
35	27	4.4	356	4	BG486009	BG486009 dab10b04.s
36	27	4.4	357	7	CP972516	CP972516 EST1045.Z
37	27	4.4	380	5	BM197491	BM197491 BM192491
38	27	4.4	382	6	CB808280	CB808280 AMGNNUC.Y
39	27	4.4	383	1	AI185107	AI185107 GES1107.X
40	27	4.4	383	1	AI388033	AI388033 GH18764.X
41	27	4.4	384	4	BI979250	BI979250 f85a08.Y
42	27	4.4	392	7	CK235610	CK235610 SB0100130
43	27	4.4	393	7	CN233362	CN233362 RJAI10C02
44	27	4.4	395	4	BM510577	BM510577 i47f09.X
45	27	4.4	398	4	BM511408	BM511408 i47f09.Y

ALIGNMENTS

RESULT 1
LOCUS CK360119 543 bp mRNA linear EST 23-DEC-2003
DEFINITION AGENCOURT 17116318 NIH_ZGC_4 Danio rerio cDNA clone IMAGE:7087988
5', mRNA sequence.
ACCESSION CK360119
VERSION CK360119.1 GI:40326051
KEYWORDS EST.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio

REFERENCE
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: c9apbs-remail.nih.gov
Tissue Procurement: John Ngai, Univ of CA, Berkeley
cDNA Library Preparation: Dr. Sumio Sugano
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LIML at:
http://image.llnl.gov
Plate: LIML4919 row: b column: 19
High quality sequence, stop: 393.
Location/Qualifiers
1..543

FEATURES
source
1..543
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7958"
/clone="IMAGE:7087988"
/lab_host="DH10B T0NA"
/clone_id="NIH_ZGC_4"
/note="Organ: brain/CNS; Vector: pXEBIS-FU3; Site_1:
Drat1r; Site_2: Drat1r"

ORIGIN

Query Match 4.7% Score 29; DB 7; Length 543;
Best Local Similarity 100.0%; Pred. No. 0.0028;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 CGACGAGAAAAA 620
 DB 282 CGACGAGAAAAA 310

RESULT 2
 LOCUS BFE61816
 DEFINITION 601 bp mRNA linear EST 21-DEC-2000
 602117624P1 Soares mammary gland_NMLMG Mus musculus cDNA clone
 IMAGE:3468916 5', mRNA sequence.

ACCESSION BFE61816
 VERSION BFE61816.1 GI:11955711
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 601)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cga@bbs-rcmail.nih.gov
 Tissue Procurement: M. Bento Soares, Ph.D.
 cDNA Library Preparation: Bento Soares and M. Fatima Bonaldo
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LAM8476 row: p column: 05
 High quality sequence start: 7
 High quality sequence stop: 479.
 Location/Qualifiers
 1..601
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:3468916"
 /sex="female (lactating)"
 /tissue_type="mammary gland"
 /lab_host="DH10B"
 /clone_lib="Soares mammary gland_NMLMG"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; 1st strand cDNA was prepared from mammary
 gland tissue from a lactating female, and was then primed
 with a Not I - oligo(dT) primer. Double-stranded cDNA was
 ligated to Eco RI adaptors (Pharmacia), digested with Not
 I and cloned into the Not I and Eco RI sites of the
 modified pT73 vector. Library is normalized. Library
 was constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN
 Query Match 4.7%; Score 29; DB 2; Length 601;
 Best Local Similarity 100.0%; Pred. No. 0.0028;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 592 CGACGAGAAAAA 620
 DB 424 CGACGAGAAAAA 452

RESULT 3
 LOCUS A1365051
 DEFINITION 199 bp mRNA linear EST 13-FEB-1999
 6121204.X1 NCI CGAP_G04 Homo sapiens cDNA clone IMAGE:1947343 3',
 mRNA sequence.

ACCESSION A1365051
 VERSION A1365051.1 GI:4124740
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 199)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-rcmail.nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 www.bio.lnl.gov/bbtp/image/image.html
 Insert length: 1431 Std Error: 0.00
 Seq primer: -40bp from G1bp
 High quality sequence stop: 189.
 Location/Qualifiers
 1..199
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1947343"
 /tissue_type="pooled germ cell tumors"
 /lab_host="DH10B"
 /clone_lib="NCI-CGAP_G04"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; 1st strand cDNA was prepared from 3 pooled
 germ cell tumors, and was then primed with a Not I -
 oligo(dT) primer. Double-stranded cDNA was ligated to Eco
 RI adaptors (Pharmacia), digested with Not I and cloned
 into the Not I and Eco RI sites of the modified pT73
 vector. Library is normalized. Library was constructed by
 Bento Soares and M. Fatima Bonaldo."

ORIGIN
 Query Match 4.5%; Score 28; DB 1; Length 199;
 Best Local Similarity 100.0%; Pred. No. 0.01;
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 593 GAACGAGAAAAA 620
 DB 48 GAACGAGAAAAA 21

RESULT 4
 LOCUS BE049016
 DEFINITION 285 bp mRNA linear EST 08-JUN-2000
 BE049016 x164f11.x2 NCI CGAP_OY26 Homo sapiens cDNA clone IMAGE:2764941 3',
 similar to contains element L1 D1 repetitive element, mRNA
 sequence.

ACCESSION BE049016
 VERSION BE049016.1 GI:8366071
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 285)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-rcmail.nih.gov
 Tissue Procurement: Monica Brown, M.D., Elise Kohn, M.D., Michael
 R. Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: David B. Krizman, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be

Tue Nov 23 09:27:42 2004

us-10-822-496-5.olig.rml

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2004, 08:50:30 ; Search time 91 Seconds
(without alignments)

4842.737 Million cell updates/sec

Title: US-10-822-496-5

Perfect score: 620

Sequence: 1 gttcaaaataactatttca.....aaaaaaaaaaaaaaaaaaaaa 620

Scoring table: OLIGO_NUC

Gapop 60.0, Gapext 60.0

Searched: 824507 seqs, 355394441 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 45 summaries

Database : Issued Patents NA:*

1: /cgn2_6/prodata/1/ina/5A.COMB.seq:*

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4.2	1384	3	US-09-227-357-142	Sequence 142, App
2	4.0	685	2	US-08-254-354-1	Sequence 1, Appl
3	4.0	685	2	US-08-254-354-3	Sequence 3, Appl
4	4.0	685	5	PCT-US95-06137-1	Sequence 1, Appl
5	4.0	685	5	PCT-US95-06137-3	Sequence 3, Appl
6	4.0	1361	4	US-09-614-912-37	Sequence 37, Appl
7	4.0	5207	4	US-09-858-664A-1	Sequence 1, Appl
8	4.0	5207	4	US-10-274-978-1	Sequence 1, Appl
9	4.0	5207	4	US-10-274-978-3	Sequence 3, Appl
10	3.9	53	4	US-09-621-976-11819	Sequence 11819, A
11	3.9	55	4	US-09-621-976-11651	Sequence 11651, A
12	3.9	57	4	US-09-513-998C-19821	Sequence 19821, A
13	3.9	59	4	US-09-621-976-11876	Sequence 11876, A
14	3.9	66	4	US-08-776-944-11	Sequence 11, Appl
15	3.9	66	4	US-09-621-976-11887	Sequence 11887, A
16	3.9	69	4	US-09-621-976-12454	Sequence 12454, A
17	3.9	70	4	US-09-621-976-12440	Sequence 12440, A
18	3.9	72	4	US-09-621-976-12390	Sequence 12390, A
19	3.9	77	4	US-09-621-976-12266	Sequence 12266, A
20	3.9	81	4	US-09-621-976-12198	Sequence 12198, A
21	3.9	81	4	US-09-621-976-13601	Sequence 13601, A
22	3.9	82	4	US-09-621-976-11689	Sequence 11689, A
23	3.9	82	4	US-09-621-976-11841	Sequence 11841, A
24	3.9	82	4	US-09-621-976-11864	Sequence 11864, A
25	3.9	82	4	US-09-621-976-11888	Sequence 11888, A
26	3.9	82	4	US-09-621-976-11944	Sequence 11944, A
27	3.9	82	4	US-09-621-976-11949	Sequence 11949, A

28	24	3.9	82	4	US-09-621-976-12015	Sequence 12015, A
29	24	3.9	82	4	US-09-621-976-12088	Sequence 12088, A
30	24	3.9	82	4	US-09-621-976-12103	Sequence 12103, A
31	24	3.9	82	4	US-09-621-976-12127	Sequence 12127, A
32	24	3.9	82	4	US-09-621-976-12131	Sequence 12131, A
33	24	3.9	82	4	US-09-621-976-12137	Sequence 12137, A
34	24	3.9	82	4	US-09-621-976-12322	Sequence 12322, A
35	24	3.9	82	4	US-09-621-976-12368	Sequence 12368, A
36	24	3.9	82	4	US-09-621-976-13499	Sequence 13499, A
37	24	3.9	82	4	US-09-621-976-13526	Sequence 13526, A
38	24	3.9	83	4	US-09-621-976-12087	Sequence 12087, A
39	24	3.9	83	4	US-09-621-976-12175	Sequence 12175, A
40	24	3.9	83	4	US-09-621-976-12195	Sequence 12195, A
41	24	3.9	83	4	US-09-621-976-12429	Sequence 12429, A
42	24	3.9	84	4	US-09-621-976-12450	Sequence 12450, A
43	24	3.9	84	4	US-09-621-976-11710	Sequence 11710, A
44	24	3.9	91	4	US-09-621-976-12161	Sequence 12161, A
45	24	3.9	98	4	US-09-621-976-11744	Sequence 11744, A

ALIGNMENTS

RESULT 1
US-09-227-357-142
; Sequence 142, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT APPLICATION NUMBER: US/09/227,357
; EARLIER FILING DATE: 1999-01-08
; EARLIER APPLICATION NUMBER: PCT/US98/13664
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: 60/051,926
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,793
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,925
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,929
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,803
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,732
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,931
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,932
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,916
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,930
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,918
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,920
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,733
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,795
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,919
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,928
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/055,722
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,723
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,948
; EARLIER FILING DATE: 1997-08-18

EARLIER APPLICATION NUMBER: 60/055,949
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,953
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,950
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,947
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,964
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/056,360
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,684
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,984
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,954
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/058,785
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,664
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,660
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,661
EARLIER FILING DATE: 1997-09-12
NUMBER OF SEQ ID NOS: 672
SOFTWARE: Patentn Ver. 2.0
SEQ ID NO 142
LENGTH: 1384
TYPE: DNA
ORGANISM: Homo sapiens
US-09-227-357-142

Query Match 4.2% Score 26; DB 3; Length 1384;
Best Local Similarity 100.0%; Pred. No. 0.0079;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 595 ACAGAGAAAAAAAAAAAAAAAAAAAA 620
DB 1336 ACAGAGAAAAAAAAAAAAAAAAAAAA 1361

RESULT 2
US-08-254-354-1
Sequence 1, Application US/08254354
Patent No. 5952210
GENERAL INFORMATION:
APPLICANT: Creely, David P.
APPLICANT: Hauser, Scott D.
APPLICANT: Welisch, Dean J.
TITLE OF INVENTION: Nucleic Acids and Expression Vectors
TITLE OF INVENTION: Encoding Human Leukotriene C4 Synthase, Antisense
TITLE OF INVENTION: Oligonucleotides, and Methods of Use
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSER: Scott B. Feder, G. D. Searle & Co., Corporate
STREET: P. O. Box 5110
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60680
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/254,354
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

NAME: Feder, Scott B.
REGISTRATION NUMBER: 33,129
REFERENCE/DOCKET NUMBER: 2800
TELECOMMUNICATION INFORMATION:
TELEPHONE: (708) 470-6501
TELEFAX: (708) 470-6881
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 685 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 87..539
US-08-254-354-1

Query Match 4.0% Score 25; DB 2; Length 685;
Best Local Similarity 100.0%; Pred. No. 0.024;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 596 CGAGAAAAAAAAAAAAAAAAAAAA 620
DB 653 CGAGAAAAAAAAAAAAAAAAAAAA 677

RESULT 3
US-08-254-354-3/C
Sequence 3, Application US/08254354
Patent No. 5952210
GENERAL INFORMATION:
APPLICANT: Creely, David P.
APPLICANT: Hauser, Scott D.
APPLICANT: Welisch, Dean J.
TITLE OF INVENTION: Nucleic Acids and Expression Vectors
TITLE OF INVENTION: Encoding Human Leukotriene C4 Synthase, Antisense
TITLE OF INVENTION: Oligonucleotides, and Methods of Use
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSER: Scott B. Feder, G. D. Searle & Co., Corporate
STREET: P. O. Box 5110
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60680
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/254,354
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Feder, Scott B.
REGISTRATION NUMBER: 33,129
REFERENCE/DOCKET NUMBER: 2800
TELECOMMUNICATION INFORMATION:
TELEPHONE: (708) 470-6501
TELEFAX: (708) 470-6881
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 685 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
ANTI-SENSE: YES
US-08-254-354-3

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 20, 2004, 08:42:52; Search time 422 Seconds

(without alignments)
7712.420 Million cell updates/sec

Title: US-10-822-496-5

Perfect score: 620
1 gttcaaaataacatttca.....aaaaaaaaaaaaaaaaaaaaa 620

Scoring table: OLIGO_NUC
Gapop 60.0, Gapext 60.0

Searched: 4134886 seqs, 2624710521 residues

Word size: 0

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

N_Geneseq_23sep04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001s:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	27	4.4	103	ADCI7856	Adci7856 Monobactam
2	27	4.4	165	AB273026	AB273026 Rice Leaf
3	27	4.4	251	AA213978	AA213978 Human Gen
4	27	4.4	473	ABV95307	ABV95307 Human Pan
5	26	4.2	120	AA895676	AA895676 Exo20 nuc
6	26	4.2	157	AB186076	AB186076 Human cya
7	26	4.2	341	AA095361	AA095361 Cat flea
8	26	4.2	383	AA119639	AA119639 Human bre
9	26	4.2	792	AA194836	AA194836 Human neu
10	26	4.2	1054	AA487440	AA487440 Thraustoc
11	26	4.2	1384	AA27442	AA27442 Human sec
12	26	4.2	1384	ADA07321	ADA07321 Human DNA
13	26	4.2	1384	ADN41015	ADN41015 Novel hum
14	26	4.2	1462	ADN40910	ADN40910 Novel hum
15	26	4.2	1696	AB073684	AB073684 Human pot
16	26	4.2	2617	AA170851	AA170851 Human hep
17	26	4.2	11757	AAK65731	AAK65731 Human imm
18	26	4.0	325	ABV38209	ABV38209 Human pro
19	25	4.0	330	ACH31895	ACH31895 Human end
20	25	4.0	331	ACH36401	ACH36401 Human end
21	25	4.0	394	AA110059	AA110059 Human bre

22	25	4.0	467	ABV58112	ABV58112 Human pro
23	25	4.0	520	ABV56379	ABV56379 Human pro
24	25	4.0	587	AAH25771	AAH25771 Human cel
25	25	4.0	587	AB234920	AB234920 Human gen
26	25	4.0	587	ADP47748	ADP47748 CXB encod
27	25	4.0	603	AAH25770	AAH25770 Human cel
28	25	4.0	603	ADP47746	ADP47746 CK-HA enc
29	25	4.0	685	AA11632	AA11632 Leukotrin
30	25	4.0	685	AA11631	AA11631 Human leu
31	25	4.0	1027	AA199518	AA199518 Human pol
32	25	4.0	1034	AA199531	AA199531 Human pol
33	25	4.0	1361	ADN94237	ADN94237 Rice GA-2
34	25	4.0	1743	ABAO0091	ABAO0091 Breast/Co
35	25	4.0	1876	ACC43639	ACC43639 Nucleotid
36	25	4.0	2700	AAH33268	AAH33268 Human col
37	25	4.0	4175	AA062285	AA062285 CDNA encod
38	25	4.0	5207	AA143909	AA143909 Human kin
39	25	4.0	5207	AA143908	AA143908 Human kin
40	25	4.0	8106	ABX11641	ABX11641 Human ser
41	25	4.0	24120	ABX11642	ABX11642 Human ser
42	25	4.0	32134	AA199172	AA199172 Human exc
43	25	4.0	32134	AA163522	AA163522 Human kid
44	25	4.0	32192	AA199173	AA199173 Human exc
45	25	4.0	32192	AA163523	AA163523 Human kid

ALIGNMENTS

RESULT 1	ADCI7856	ADCI7856 standard; RNA; 103 BP.
ID	ADCI7856	
XX	XX	
AC	ADCI7856;	
XX	XX	
DT	18-DEC-2003	(first entry)
XX	XX	
DE	Monobactam related 81 biocatalyst subpopulation consensus SEQ ID NO:106.	
XX	XX	
KW	ss; monobactam; antibacterial; PBP2a; inhibitor;	
KW	methicillin resistant Staphylococcus aureus; MRSA; lactam antibiotic.	
XX	XX	
OS	Synthetic.	
XX	XX	
FT	Key	Location/Qualifiers
FT	modified_base	1..103
XX	FT	/*tag= a
XX	XX	/mod_base= 5-(4-pyridylmethyl)U
PN	WO2003051314-A2.	
XX	XX	
PD	26-JUN-2003.	
XX	XX	
PF	18-DEC-2002; 2002WO-US040739.	
XX	XX	
PR	18-DEC-2001; 2001US-0340255P.	
XX	XX	
PA	(INVE-) INVENTUX INC.	
XX	XX	
PI	Eaton B, Tarasow T, Nieuwlandt D, Dewey T;	
XX	XX	
DR	WPI; 2003-618003/58.	
XX	XX	
PT	New monobactam compounds used as antibacterial agents against e.g.	
XX	XX	methicillin resistant Staphylococcus aureus.
PS	Disclosure; SEQ ID NO 106; 64bp; English.	
XX	XX	
CC	The invention relates to novel monobactam compounds. A compound of the	
CC	invention has antibacterial activity, and acts as a PBP2a inhibitor. The	
CC	compounds are used as antibacterial agents. The monobactam compounds	
CC	restore sensitivity of methicillin resistant Staphylococcus aureus to	
CC	lactam antibiotic by targeting the molecular mechanism of resistance. The	

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2004, 15:32:53, Search time 2977 Seconds

(without alignments)
7589.061 Million cell updates/sec

Title: US-10-822-496-5

Sequence: 1 gtctcaataatacattttaa.....aaaaaaaaaaaaaaaaaaaa 620

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database:

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	59.6	9.6	997	CNS0057E	AL060767 Drosophila
C 2	59.4	9.6	987	CNS00418	AL066537 Drosophila
C 3	57	9.2	536	AG690430	AG690430 rdxp0082E
C 4	57	9.2	1263	CG745034	CG745034 P037-4-C0
C 5	56.6	9.1	831	AG301159	AG301159 Mus muscu
C 6	56.4	9.1	957	CNS0060US	AL065591 Drosophila
C 7	55.8	8.0	1451	AG393206	AG393206 Mus muscu
C 8	55.4	8.9	1201	CNS0108P	AL098659 Drosophila
C 9	55.2	8.9	839	AG372305	AG372305 Mus muscu
C 10	55	8.9	809	AG473719	AG473719 Mus muscu
C 11	54.8	8.8	1100	CNS000EX	AL069107 Drosophila
C 12	54.6	8.8	1209	CG749482	CG749482 P043-3-H1
C 13	54.6	8.8	793	AG469170	AG469170 Mus muscu
C 14	54.6	8.8	806	AG344150	AG344150 Mus muscu
C 15	54.6	8.8	865	AG376220	AG376220 Mus muscu
C 16	54.6	8.8	959	CNS008WZ	AL052079 Drosophila
C 17	54.2	8.7	851	AG506735	AG506735 Mus muscu
C 18	54	8.7	786	AG51082	AG51082 Mus muscu
C 19	53.8	8.7	391	AG954378	AG954378 rdxp0070D
C 20	53.8	8.7	901	CNS026A5	AL183110 Tetradon
C 21	53.4	8.6	1101	CNS00F3N	AL069996 Drosophila
C 22	53.2	8.6	878	CG926174	CG926174 t086j02ba
C 23	53	8.5	760	AG534996	AG534996 Mus muscu
C 24	53	8.5	778	AG474409	AG474409 Mus muscu

25	53	8.5	885	9	CNS018H7	AL109333 Drosophila
C 26	53	8.5	934	4	BM415736	BM415736 OP20817 M
C 27	53	8.5	1027	9	AG523385	AG523385 Mus muscu
C 28	52.8	8.5	1223	9	CNS015ZH	AL106103 Drosophila
C 29	52.8	8.5	1461	9	AG141873	AG141873 Pan trogl
C 30	52.8	8.5	1969	3	CG670394	CG670394 Tetradon
C 31	52.6	8.5	255	2	BS993423	BS993423 Mus muscu
C 32	52.6	8.5	766	9	AG601460	AG601460 Mus muscu
C 33	52.4	8.5	756	9	AG429320	AG429320 Mus muscu
C 34	52.4	8.5	854	9	AG523762	AG523762 Mus muscu
C 35	52.4	8.5	974	9	CNS049SW	AL128098 Tetradon
C 36	52.4	8.5	1023	5	BU160131	BU160131 AGENCOURT
C 37	52.4	8.5	1108	9	CU030161	CU030161 CH216-30C
C 38	52.4	8.5	1239	9	CG750217	CG750217 P044-3-H0
C 39	52.4	8.5	1328	9	CG746913	CG746913 P040-2-C0
C 40	52.2	8.4	756	8	B21789	B21789 F6P1-SP6 IG
C 41	52.2	8.4	813	9	AG328054	AG328054 Mus muscu
C 42	52.2	8.4	819	9	AG361489	AG361489 Mus muscu
C 43	52.2	8.4	1053	4	BM415599	BM415599 OP20676 M
C 44	52.2	8.4	1101	9	CNS0170G	AL108298 Drosophila
C 45	52.2	8.4	1157	9	CU076711	CU076711 CH216-152

ALIGNMENTS

RESULT 1
CNS0057E 997 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TERT end of BAC #
DEFINITION BAC12K22 of RPCL-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

ACCESSION AL060767.1 GI:4943573
VERSION AL060767
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
Web : www.genoscope.cns.fr)

COMMENT

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazuhiro Osoegawa and Aaron Mammose in Peter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCL-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2, cn bw sp, the same strain used for the BDGP's F1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BAC PAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source
1. 997
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BAC12K22"
/clone_lib="RPCL-98"
/note="end : TERT"

ORIGIN

Query Match 9.6%; Score 59.6; DB 9; Length 997;
Best Local Similarity 24.7%; Pred. No. 7.4e-05;

Matches	41; Conservative	78; Mismatches	47; Indels	0; Gaps
OY	455	GAATGACGACCAAAAGCAATGTGGACAGCTGATATGGAGCGCGCGGCGGAGAGC	514	
Db	814	RAKRRRRRAAARARARARARRRRRARAGRRRRRGGRRRRRRRRRRRRARARAR	755	
OY	515	AGAGAAAAGGCAACAATAGAGGCGAGATTAAGGGGGGAGGCGCAAGATTAACGAACT	574	
Db	754	ARAGARAAAARRRRRAGAAARRRRRRRRRRRRRRRRRRRRRRRRARARARAGAAARR	695	
OY	575	GCAGTGGGAGAAAACCTCGAACGAAAAAATTTTTTTTTTTTTTTTTTTTTTTT	620	
Db	694	RRRRRRGGAGARRRRGRRRRRGGAGARRRRRRRRRRRRRRRRRRRRRRRR	649	
RESULT 2				
CNS00418/c				
LOCUS				
DEFINITION	CNS00418	987 bp	DNA	linear
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
COMMENT				
FEATURES				
SOURCE				
ORIGIN				
Query Match	9.6%;	Score 59.4;	DB 9;	Length 987;
Best Local Similarity	32.3%;	Pred. No. 8.4e-05;		
Matches 54;	Conservative 61;	Mismatches 52;	Indels 0;	Gaps 0;
OY	454	GGATGAGNCGACGAAAGACAAATGTGGACAGCTGATATGGAGCGCGCGGGAAGT	513	
Db	878	RGAGRARRRGRRGRR	819	
OY	514	GAGAGAAAAGGCAACAATAGAGGCGAGATTAAGGGGGGAGGCGCAAGAAATTAACAGAAC	573	
Db	818	GGRRRAAARRRRRAGARARRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR	759	
OY	574	TGCGATGGGAGAAAACGTGCAACGAAAAAATTTTTTTTTTTTTTTTTTTTTTTT	620	

	Db	758	GAGGAGAAGACGAATTAATAAAGGCGGCAAGAACTTAAACAAGAAC	712
	RESULT 3			
	A0690430/c	536 bp	DNA	linear GSS 01-JUL-1999
	LOCUS	nbxb0082E17r	CUGI Rice BAC Library	Oryza sativa (japonica
	DEFINITION	cultivar-group)	genomic clone nbxb0082E17r,	genomic survey
	sequence.			
	A0690430			
	VERSION	A0690430.1	GI:5331598	GSS.
	KEYWORDS			
	SOURCE	Oryza sativa (japonica cultivar-group)		
	ORGANISM	Oryza sativa (japonica cultivar-group)		
		Burayocsa, Vitidiplantae, Streptophyta, Embryophyta; Tracheophyta;		
		Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;		
		Eriactoidae; Oryzaceae; Oryza.		
		1 (bases 1 to 536)		
	REFERENCE	Wing, R.A. and Dean, R.A.		
	AUTHORS	A BAC End Sequencing Framework to Sequence the Rice Genome		
	TITLE	Unpublished (1998)		
	JOURNAL	Contact: Wing RA		
	COMMENT	Clemson University Genomics Institute		
		100 Jordan Hall, Clemson, SC 29634, USA		
		Tel.: 864 656 7298		
		Fax: 864 656 4293		
		Email: twing@clemson.edu		
		Seq primer: GSAAACAGCTATGACCATG		
		Class: BAC ends		
	FEATURES	High quality sequence stop: 154.		
	source	location/Qualifiers		
		1..536		
		/organism="Oryza sativa (japonica cultivar-group)"		
		/mol_type="Genomic DNA"		
		/cultivar="Nippobare"		
		/db_xref="taxon:39947"		
		/clone="nbxb0082E17r"		
		/tissue_type="leaf"		
		/lab_host="E. coli DH10B"		
		/lab_host="CUGI Rice BAC Library"		
		/note="(vector: pbeloBAC11; site_1: HindIII; site_2:		
		HindIII; Rice is one of two most popular grains in the		
		world. Half of the world population especially those		
		inhabiting highly populated areas of the humid tropics		
		and subtropics, rely on rice as their primary source of		
		carbohydrate. Monocotyledonous rice is a diploid plant		
		(2n=24) with a haploid genome equivalent of 431 Mbp		
		(Arumuganathan and Earle, 1991). The relatively small		
		genome of rice, three times larger than that of		
		Arabidopsis, makes it suitable for genomic studies. In		
		order to facilitate positional cloning, physical mapping		
		and genome sequencing of rice, we have constructed a BAC		
		library from Oryza sativa, Nippobare variety. The		
		library contains 36,864 clones with an average insert size		
		of 18.5 Kb providing 10.9 haploid genome equivalents. The		
		deep coverage allows the isolation a particular sequence		
		with a probability of 99.9%. Two high density filters,		
		each containing 18,432 clones (doubly spotted), represent		
		the whole library for colony screening."		
	ORIGIN			
	Query Match	9.2%; Score 57; DB 8; Length 536;		
	Best Local Similarity	57.5%; Pred. No. 0.00033;		
	Matches	96; Conservative 0; Mismatches 71; Indels 0; Gaps 0;		
	454	GGAATGGAACACAGAAAACAATGTGGGACACTGATTGAGGGCGAGCGGAGAGT	513	
	443	CGAAGCGCGGAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAGAAAAG	384	
	514	GACGAAAAGGCGAACAATATAGGCGCAATTAAGGCGGCGGCGCAAGAACTTA	573	

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2004, 15:37:28 ; Search time 90 Seconds

(without alignments)
4896.546 Million cell updates/sec

Title: US-10-822-496-5

Perfect score: 620
Sequence: 1 gtcacaaataacattttaa.....aaaaaaaaaaaaaaaaaaaaa 620

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: Issued Patents NA:
2: /cgn2_6/prodata/1/ina/5A COMB.seq:*
3: /cgn2_6/prodata/1/ina/5B COMB.seq:*
4: /cgn2_6/prodata/1/ina/5C COMB.seq:*
5: /cgn2_6/prodata/1/ina/5D COMB.seq:*
6: /cgn2_6/prodata/1/ina/5E COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	68	11.0	7218	1	Sequence 14, Appl
2	52.6	8.5	289	3	Sequence 17, Appl
3	52.6	8.5	289	3	Sequence 17, Appl
4	48	7.7	277	3	Sequence 3, Appl
5	48	7.7	277	3	Sequence 3, Appl
6	43.4	7.0	893	4	Sequence 46, Appl
7	42.6	6.9	123	3	Sequence 1, Appl
8	42.6	6.9	123	3	Sequence 1, Appl
9	42.2	6.8	261	4	Sequence 466, App
10	42.2	6.8	1923	4	Sequence 466, App
11	41.2	6.6	705	4	Sequence 5061, App
12	41.2	6.6	705	4	Sequence 2043, App
13	40.8	6.6	372	4	Sequence 13, Appl
14	40.8	6.6	372	4	Sequence 13, Appl
15	40.6	6.5	43795	3	Sequence 101, App
16	40.2	6.5	16442	3	Sequence 208, App
17	40.2	6.5	16442	3	Sequence 208, App
18	39.4	6.4	356	4	Sequence 22, Appl
19	39.4	6.4	356	4	Sequence 22, Appl
20	39.4	6.4	2327	4	Sequence 20, Appl
21	39.4	6.4	2674	4	Sequence 19, Appl
22	39.4	6.4	2771	4	Sequence 18, Appl
23	39.4	6.4	5860	4	Sequence 17, Appl
24	39.4	6.4	9646	3	Sequence 1, Appl
25	39.4	6.4	9646	3	Sequence 1, Appl
26	39.4	6.4	12980	3	Sequence 5, Appl
27	39.4	6.4	12980	3	Sequence 5, Appl

28	39.2	6.3	270	2	US-08-520-678A-30	Sequence 30, Appl
29	39.2	6.3	270	3	US-08-897-126-30	Sequence 30, Appl
30	39.2	6.3	53526	3	US-08-658-136-2	Sequence 2, Appl
31	39.2	6.3	53577	4	US-08-658-136-1	Sequence 1, Appl
32	38.8	6.3	548	4	US-09-621-976-894	Sequence 894, App
33	38.8	6.3	696	3	US-09-461-697-193	Sequence 193, App
34	38.8	6.3	699	3	US-09-461-697-191	Sequence 191, App
35	38.8	6.3	717	3	US-09-461-697-189	Sequence 189, App
36	38.8	6.3	774	3	US-09-461-697-187	Sequence 187, App
37	38.8	6.3	819	3	US-09-461-697-185	Sequence 185, App
38	38.6	6.3	1669	3	US-09-461-697-184	Sequence 184, App
39	38.6	6.2	168174	4	US-10-071-411A-63	Sequence 63, Appl
40	38.6	6.2	168273	4	US-10-071-411A-2	Sequence 2, Appl
41	38.4	6.2	260	2	US-08-520-678A-29	Sequence 29, Appl
42	38.4	6.2	260	3	US-08-897-126-29	Sequence 29, Appl
43	38.4	6.2	8642	4	US-10-029-907-2	Sequence 2, Appl
44	38.2	6.2	376	2	US-08-623-906A-18	Sequence 18, Appl
45	38.2	6.2	669	4	US-09-669-751-79	Sequence 79, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/C
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHIEFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29, 766
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZapc-F15
US-08-232-463-14
Query Match 11.0%; Score 68; DB 1; Length 7218;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2004, 10:03:27 ; Search time 404 Seconds

(without alignments)
8056.042 Million cell updates/sec

Title: US-10-822-496-5

Perfect score: 620
Sequence: 1 gttcaaaataacattttaa.....aaaaaaaaaaaaaaaaaaaaa 620

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 413486 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_23Sep04:*

1: geneseq1980s:*\n2: geneseq1990s:*\n3: geneseq2000s:*\n4: geneseq2001as:*\n5: geneseq2001bs:*\n6: geneseq2002as:*\n7: geneseq2002bs:*\n8: geneseq2003as:*\n9: geneseq2003bs:*\n10: geneseq2003cs:*\n11: geneseq2003ds:*\n12: geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	47.8	7.7	1840	10	ADc87550 Human GPC
2	47.6	7.7	511	10	ADb59059 Toxicity
3	47.6	7.7	511	10	ADb53812 Primary x
4	47.4	7.6	409	4	AA184865 Human pol
5	47	7.6	1431	12	ADQ25488 Human sof
6	46.6	7.5	2502	5	AA87937 DNA encod
7	46.2	7.5	210	6	AB186389 Human ova
8	46.2	7.5	1563	6	ABL53204 Nucleoid
9	46.2	7.5	11303	8	AA154213 SR protei
C 10	46	7.4	1426	10	ADc87528 Human GPC
11	45.8	7.4	498	5	AA869552 DNA encod
12	45.8	7.4	4238	10	ADb53355 Primary x
C 13	45.4	7.3	5059	6	AA84332 Stealth v
14	45.4	7.3	220895	6	ABK84798 Human cDN
15	45.4	7.3	312477	12	ADPe9744 Human ROC
16	44.6	7.2	654	5	AA868925 DNA encod
17	44.4	7.2	299	4	AA124183 Probe #14
18	44.4	7.2	299	4	ABa63302 Human foe
19	44.4	7.2	299	4	AA149470 Probe #18
20	44.4	7.2	299	4	ABa51297 Human bre
21	44.4	7.2	299	4	ABa36233 Probe #14

22	44.4	7.2	299	4	AAK43403 Human bora
23	44.4	7.2	299	4	AAK17588 Human liv
24	44.4	7.2	299	4	ABa43023 Human liv
25	44.4	7.2	299	5	AA109746 Probe #97
26	44.4	7.2	299	6	ABs17496 Human gen
27	44.4	7.2	12700	2	AAV62133 HSV-2 str
C 28	44.2	7.1	2350	10	ADc87384 Human GPC
29	44	7.0	893	2	AAx57416 Rat gene
30	43.4	7.0	893	12	ADJ74215 Rat cDNA
31	43.4	7.0	264	5	ABV58299 Human pro
32	43.2	7.0	265	5	ABV58274 Human pro
33	43.2	7.0	265	4	AA116963 Probe #68
C 34	43.2	7.0	598	4	AAa61071 Human foe
C 35	43.2	7.0	598	4	AA140968 Probe #96
C 36	43.2	7.0	598	4	ABa28973 Probe #74
C 37	43.2	7.0	598	4	AAK35253 Human bora
C 38	43.2	7.0	598	4	AAK09364 Human liv
C 39	43.2	7.0	598	6	ABs09667 Human gen
C 40	43.2	7.0	598	12	ADb53608 Novel mou
C 41	43.2	7.0	1109	12	ADP81772 Human MD-
C 42	43.2	7.0	67674	12	AA869541 DNA encod
C 43	43.2	6.9	390	5	AA869541 DNA encod
C 44	43	6.9	206	5	ADL44537 Human ova
C 45	42.8	6.9	206	5	ADL44537 Human ova

ALIGNMENTS

RESULT 1
ADc87550/c
ID ADc87550 standard; DNA; 1840 BP.
XX
AC ADc87550:
XX
DT 01-JAN-2004 (first entry)
XX
DE Human GPCR gene SEQ ID NO:2003.
XX
KM db; gene; human; GPCR;
XX
KW guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1270724-A2.
XX
PD 02-JAN-2003.
XX
PF 18-JUN-2002; 2002EP-00013517.
XX
PR 18-JUN-2001; 2001JP-00246789.
XX
PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI Suwa M, Asei K, Akiyama Y, Aburatani H;
XX
DR WPI; 2003-315783/31.
XX
PT New polynucleotide, useful for preparing a composition for treating a
PT patient in need of increased or suppressed activity or expression of the
PT guanosine triphosphate-binding protein coupled receptor.
XX
PS Claim 1; SEQ ID NO 2003; 28pp; English.
XX
XX The invention relates to a novel polynucleotide encoding a guanosine
XX triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
XX the invention may have a use in gene therapy. The polynucleotide and
XX polypeptide are useful for preparing a composition for treating a patient
XX in need of increased or suppressed activity or expression of the
XX guanosine triphosphate-binding protein coupled receptor. The
XX polynucleotide sequences shown in ADc8548-ADc87616 encode GPCR's of the

CC Invention.
XX Sequence 1840 BP; 58 A; 537 C; 153 G; 768 T; 0 U; 324 Other;
SQ Query Match 7.7%; Score 47.8; DB 10; Length 1840;
Best Local Similarity 54.5%; Pred. No. 0.0032;
Matches 91; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
CC 454 GGAATGAGACCGAAGACATGTGGACAGCTGATATGAGAGCGGCGGGAAT 513
DB GAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 761
CC 514 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 573
DB GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 701
CC 574 TGCACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 620
DB 700 AAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 654
RESULT 2
ADBS9059 standard; DNA; 511 BP.
AC ADB59059;
DT 04-DEC-2003 (first entry)
DE Toxicity-related gene, SEQ ID 4085.
XX Toxic; toxin; gene expression profile; hepatotoxicity; liver;
KW drug screening; toxicity assay; ds.
OS Unidentified.
XX WC2003064624-A2.
PN 07-AUG-2003.
PD 31-JAN-2003; 2003WO-US003194.
PE 31-JAN-2002; 2002US-00060087.
PR 15-MAR-2002; 2002US-0364045P.
PR 15-MAR-2002; 2002US-0364055P.
PR 30-DEC-2002; 2002US-0436643P.
XX (GENE-) GENE LOGIC INC.
PA Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M,
PI WPI; 2003-689530/65.
DR WPI; 2003-689530/65.
XX Predicting a toxic effect of a compound, useful in identifying toxicity
PT markers in liver tissues or cells for drug screening and toxicity assays,
PT comprises preparing gene expression profile of tissue or cells exposed to
the compound.
PS Claim 1; SEQ ID NO 4085; 1156bp; English.
XX The present invention relates to a method for predicting a toxic effect
CC of a tissue or cell sample exposed to the compound, and comparing the
CC gene expression profile to a database comprising SEQ ID 1-4925, where
CC differential expression of the gene indicates at least one toxic effect.
CC The method is useful for predicting at least one toxic effect of a
CC compound, predicting hepatotoxicity or the progression of a toxic effect
CC of a compound, identifying an agent that modulates the onset or
CC progression of a toxic response, predicting the cellular pathways that a
CC compound modulates in a cell, and identifying an agent that modulates at
CC least one activity of a protein. The method and compositions of the
CC present invention using a database of genes having liver toxin-induced
CC differential expression, are useful in identifying toxicity markers in

CC liver tissues or cells for drug screening and toxicity assays. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pat_sequences.
XX Sequence 511 BP; 236 A; 46 C; 168 G; 61 T; 0 U; 0 Other;
SQ Query Match 7.7%; Score 47.6; DB 10; Length 511;
Best Local Similarity 55.4%; Pred. No. 0.0021;
Matches 92; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
CC 455 GGAATGAGACCGAAGACATGTGGACAGCTGATATGAGAGCGGCGGGAATG 514
DB GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 370
CC 515 AGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 574
DB GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 430
CC 575 GCACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 620
DB 431 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 476
RESULT 3
ADBS3812 standard; DNA; 511 BP.
AC ADB53812;
DT 04-DEC-2003 (first entry)
DE Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:4354.
XX toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;
KW toxicity marker; toxicity progression; drug screening;
KW primary rat hepatocyte toxicity modelling; gene; ds.
OS Rattus norvegicus.
XX WC2003065993-A2.
PN 14-AUG-2003.
PD 04-FEB-2003; 2003WO-US003482.
PE 04-FEB-2002; 2002US-0353171P.
PR 13-MAR-2002; 2002US-0363534P.
PR 08-APR-2002; 2002US-0370248P.
PR 10-APR-2002; 2002US-0371134P.
PR 10-APR-2002; 2002US-0371135P.
PR 11-APR-2002; 2002US-0371150P.
PR 19-APR-2002; 2002US-0373601P.
PR 19-APR-2002; 2002US-0373602P.
PR 22-APR-2002; 2002US-0374139P.
PR 08-MAY-2002; 2002US-0378522P.
PR 09-MAY-2002; 2002US-0378653P.
PR 09-MAY-2002; 2002US-0378655P.
PR 09-JUL-2002; 2002US-0394230P.
PR 09-JUL-2002; 2002US-0394253P.
PR 04-SEP-2002; 2002US-0407688P.
PR 28-JAN-2003; 2003US-0442900P.
XX (GENE-) GENE LOGIC INC.
PA Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Orr M,
PI Elashoff M,
PI WPI; 2003-731472/69.
DR WPI; 2003-731472/69.
XX Determining if a compound induces a toxic effect on a tissue or cell, for

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: November 16, 2004, 13:33:03 / Search time 36 Seconds
(without alignments)
3314.133 Million cell updates/sec

Title: US-10-822-496-5

Perfect score: 201
Sequence: 1 gttcaaaataacattttaa.....aaaaaaaaaaaaaaaaaaaa 620

Scoring table:

OLIGO
Xgapop 60.0, Xgapext 60.0
Ygapop 60.0, Ygapext 60.0
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 283416 segs, 96216763 residues

Word size: 1

Total number of hits satisfying chosen parameters: 565918

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Command line parameters:
-MODE=frame_n2p.model -DEV=xlp
-O/CGN2.1/USFTC.spool.p/US10822496/runat.16112004.100157.24608/app.query.fasta.1.775
-DB=PIR_75 -QFMT=fastaan -SUFRTX=0.1112p.rpr -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=0.1112p.rpr -TRANS=human40.cdi -LIST=45
-DOCALLGN=200 -THR.SCORE=quality -THR.MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10822496.CGN.1.1.63 @runat.16112004.100157.24608 -NCPU=6 -ICPU=3
-NO WAMP -LARGEQUERY -NEG.SCORES=0 -WAIT -DSELOCK=100 -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -Fgapop=6
-Fgapext=7 -YGAPOP=60 -YGAPEXT=60 -DELop=6 -DELExt=7

Database:

PIR_75: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	9	4.4	106	2	A71615
C 2	9	4.4	111	2	S64364
C 3	9	4.5	280	2	T34540
C 4	9	4.5	784	2	A33633
C 5	9	4.5	2010	2	B71616
C 6	8	4.0	46	2	T28799
C 7	8	3.9	107	2	T29406
C 8	8	4.0	127	2	A41061
C 9	8	4.0	130	2	S55141
C 10	8	3.9	139	2	S61040
C 11	8	4.0	167	2	S38112
C 12	8	4.0	175	2	T05669
C 13	8	4.0	202	2	T11744
C 14	8	4.0	215	2	I52523

15	8	4.0	244	2	T24631	hypothetical prote
16	8	4.0	246	1	S62655	hypothetical prote
17	8	4.0	246	1	S69770	hypothetical prote
18	8	4.0	249	2	T49239	hypothetical prote
C 19	8	3.9	281	2	J06528	31k major protein,
C 20	8	3.9	282	2	H70006	N-acetylmuramoyl-L
C 21	8	4.0	295	2	S61039	hypothetical prote
C 22	8	3.9	299	2	AB1776	hypothetical prote
C 23	8	4.0	312	2	J05962	paired-box contain
C 24	8	3.9	316	2	S25843	paired-box contain
C 25	8	4.0	323	2	G83291	bi-functional cycla
C 26	8	4.0	325	2	G83291	probable transcrip
C 27	8	4.0	335	2	S5961	paired-box contain
C 28	8	4.0	349	2	J05827	paired-box contain
C 29	8	4.0	350	2	U0202	paired-box contain
C 30	8	4.0	357	2	J05828	paired-box contain
C 31	8	3.9	358	2	E70333	hypothetical prote
C 32	8	4.0	362	2	T15369	hypothetical prote
C 33	8	4.0	362	2	A13288	phosphoribosylamin
C 34	8	4.0	392	2	T09371	hypothetical prote
C 35	8	3.9	401	2	T02958	ribulose-bisphosph
C 36	8	4.0	404	2	A75192	hypothetical prote
C 37	8	4.0	419	2	B71224	hypothetical prote
C 38	8	4.0	426	2	T39431	hypothetical prote
C 39	8	4.0	467	2	T21690	probable methionin
C 40	8	4.0	469	2	E43319	hypothetical prote
C 41	8	3.9	480	2	S57842	carbon-monoxide de
C 42	8	3.9	535	2	D56680	acute myeloid leuk
C 43	8	4.0	558	2	G85430	hypothetical prote
C 44	8	4.0	560	2	T05278	hypothetical prote
C 45	8	4.0	561	2	H86442	unknown protein [1

ALIGNMENTS

RESULT 1
A71615
probable integral membrane protein PFB0415C - malaria parasite (Plasmodium falciparum)
C/Species: Plasmodium falciparum
C/Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 09-Jul-2004
C/Accession: A71615
R/Gardner, M.J.; Tettein, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Xoonin, E.V.;
.; Petrea, M.; Salberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.;
Science 282, 1126-1132, 1998
A/Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A/Reference number: A71600; MUID:99021743; PMID:9804551
A/Accession: A71615
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Residues: 1-106 <GAR>
A/Cross-References: UNIPROT:O96177; GB:AE001394; GB:AE001362; NID:93945178; PIDN:AA07187;
A/Experimental source: clone 3D7
C/Genetics:
A/Gene: PFB0415C

Alignment Scores:
Pred. No.: 1.47
Score: 9.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 4.41%
DB: 2
Gaps: 0

US-10-822-496-5 (1-620) x A71615 (1-106)

QY	620	TTTTTTTTTTTTTTTTTTTTTCGCGT	594
DB	5	Phepnepnepnepnepnepneval	13

RESULT 2
S64364
probable membrane protein YG069w - Yeast (Saccharomyces cerevisiae)
N/Alternate names: hypothetical protein G4554

